

# **MODULATORS OF CROSS-LANGUAGE INFLUENCES IN LEARNING AND PROCESSING**

EDITED BY: Anat Prior, Tamar Degani and Zofia Wodniecka

PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication





# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88976-296-5

DOI 10.3389/978-2-88976-296-5

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [frontiersin.org/about/contact](https://frontiersin.org/about/contact)

# MODULATORS OF CROSS-LANGUAGE INFLUENCES IN LEARNING AND PROCESSING

Topic Editors:

**Anat Prior**, University of Haifa, Israel

**Tamar Degani**, University of Haifa, Israel

**Zofia Wodniecka**, Jagiellonian University, Poland

**Citation:** Prior, A., Degani, T., Wodniecka, Z., eds. (2022). Modulators of Cross-Language Influences in Learning and Processing. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-296-5

# Table of Contents

- 05 Editorial: Modulators of Cross-Language Influences in Learning and Processing**  
Tamar Degani, Anat Prior and Zofia Wodniecka
- 12 Cross-Linguistic Influence on L2 Before and After Extreme Reduction in Input: The Case of Japanese Returnee Children**  
Maki Kubota, Caroline Heycock, Antonella Sorace and Jason Rothman
- 33 Individual Chunking Ability Predicts Efficient or Shallow L2 Processing: Eye-Tracking Evidence From Multiword Units in Relative Clauses**  
Manuel F. Pulido
- 51 Costs and Benefits of Native Language Similarity for Non-native Word Learning**  
Viorica Marian, James Bartolotti, Aimee van den Berg and Sayuri Hayakawa
- 65 Equal Opportunity Interference: Both L1 and L2 Influence L3 Morpho-Syntactic Processing**  
Nawras Abbas, Tamar Degani and Anat Prior
- 80 Text Complexity Modulates Cross-Linguistic Sentence Integration in L2 Reading**  
Sibylla Leon Guerrero, Veronica Whitford, Laura Mesite and Gigi Luk
- 101 Do Cross-Language Script Differences Enable Bilinguals to Function Selectively When Speaking in One Language Alone?**  
Noriko Hoshino, Anne L. Beatty-Martinez, Christian A. Navarro-Torres and Judith F. Kroll
- 113 Cross-Language Influences in the Processing of Multiword Expressions: From a First Language to Second and Back**  
Lingli Du, Irina Elgort and Anna Siyanova-Chanturia
- 131 Hands Down: Cognate Effects Persist During Written Word Production**  
Evy Woumans, Robin Clauws and Wouter Duyck
- 138 What are the Modulators of Cross-Language Syntactic Activation During Natural Reading?**  
Naomi Vingron, Pauline Palma, Jason W. Gullifer, Veronica Whitford, Deanna Friesen, Debra Jared and Debra Titone
- 156 Cognitive and Linguistic Skills Associated With Cross-Linguistic Transfer in the Production of Oral Narratives in English as a Foreign Language by Arabic- and Hebrew-Speaking Children: Finding Common Denominators**  
Susie Russak and Elena Zaretsky
- 175 The Influence of Cross-Linguistic Similarity and Language Background on Writing to Dictation**  
Antonio Iniesta, Eleonora Rossi, M. Teresa Bajo and Daniela Paolieri
- 194 Does L2 Proficiency Impact L2-L1 Transfer While Reading L1 Collocations? Evidence From Behavioral and ERP Data**  
Agnieszka Otwinowska, Marta Marecka, Alba Casado, Joanna Durlík, Jakub Szewczyk, Marcin Opacki and Zofia Wodniecka

- 209** *Crosslinguistic Influence (CLI) of Lexical Breadth and Depth in the Vocabulary of Bilingual Kindergarten Children – A Bilingual Intervention Study*  
Minna Lipner, Sharon Armon-Lotem, Joel Walters and Carmit Altman
- 227** *Eye Movement Measures of Within-Language and Cross-Language Activation During Reading in Monolingual and Bilingual Children and Adults: A Focus on Neighborhood Density Effects*  
Veronica Whitford and Marc F. Joanisse
- 241** *Child Heritage Language Development: An Interplay Between Cross-Linguistic Influence and Language-External Factors*  
Natalia Meir and Bibi Janssen
- 258** *Modulation of Cross-Language Activation During Bilingual Auditory Word Recognition: Effects of Language Experience but Not Competing Background Noise*  
Melinda Fricke



# Editorial: Modulators of Cross-Language Influences in Learning and Processing

Tamar Degani<sup>1\*</sup>, Anat Prior<sup>2\*</sup> and Zofia Wodniecka<sup>3\*</sup>

<sup>1</sup> Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel, <sup>2</sup> Department of Learning Disabilities and Edmond J. Safra Brain Research Center for Learning Disabilities, University of Haifa, Haifa, Israel,

<sup>3</sup> Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland

**Keywords:** bilingual, cross-language influences, transfer, second language learning, multilingualism, lexicon, morphosyntax

## Editorial on the Research Topic

### Modulators of Cross-Language Influences in Learning and Processing

## INTRODUCTION

Language learning and processing should be considered in the context of speakers' prior linguistic knowledge. Research accumulated over the last few decades (Jarvis and Pavlenko, 2007) has indeed demonstrated that *cross-language influences* (CLI), also termed *transfer*, are present across different language domains, for bi- and multilinguals (Cenoz et al., 2001; Puig-Mayenco et al., 2020). Research on CLI is important for several reasons. On the theoretical front, such evidence can reveal the patterns of interconnectivity of the multilingual language system and inform models of multilingual representation and activation. Further, such research carries implications for learning and instruction, in understanding when and how CLI from prior linguistic knowledge would facilitate or hinder learning.

Despite wide agreement regarding the prevalence and importance of CLI, there is quite a lot of variability in its specific manifestations across studies. Thus, the goal of the current Research Topic is to set the stage for systematically mapping the factors that may modulate the presence and nature of CLI in learning and processing. Studies included in this Research Topic investigate CLI in children and adults, across lexicon and grammar, in beginning and advanced language users. Importantly, the studies identified and tested factors that might modulate CLI. Across the papers, the modulators examined include speaker characteristics, task demands, and item/language characteristics (see **Figure 1**), thus offering a rich and nuanced description of the factors at play. In what follows, we present the collection of studies in this Research Topic according to the language domain on which they focused (see **Table 1**), as well as outline commonalities and avenues for future research.

## Lexical Domain

Seven studies in this volume examined the way in which CLI is manifested in learning and processing of single words (Fricke; Hoshino et al.; Iniesta et al.; Lipner et al.; Marian et al.; Whitford and Joanisse; Woumans et al.). Even a quick overview of the studies reveals a rich heterogeneity of the methodological approaches; from eye tracking while reading, through oral, or written production, auditory word recognition, pair-associate learning, to a vocabulary intervention. Across these studies, several factors emerged as modulating CLI, including those that relate to the experimental task, to individual differences across participants in background language and cognition, as well as to item and language specific characteristics. CLI was observed in all the

## OPEN ACCESS

### Edited and reviewed by:

Manuel Carreiras,  
Basque Center on Cognition, Brain  
and Language, Spain

### \*Correspondence:

Tamar Degani  
tdegani@research.haifa.ac.il  
Anat Prior  
aprior@edu.haifa.ac.il  
Zofia Wodniecka  
zofia.wodniecka@uj.edu.pl

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

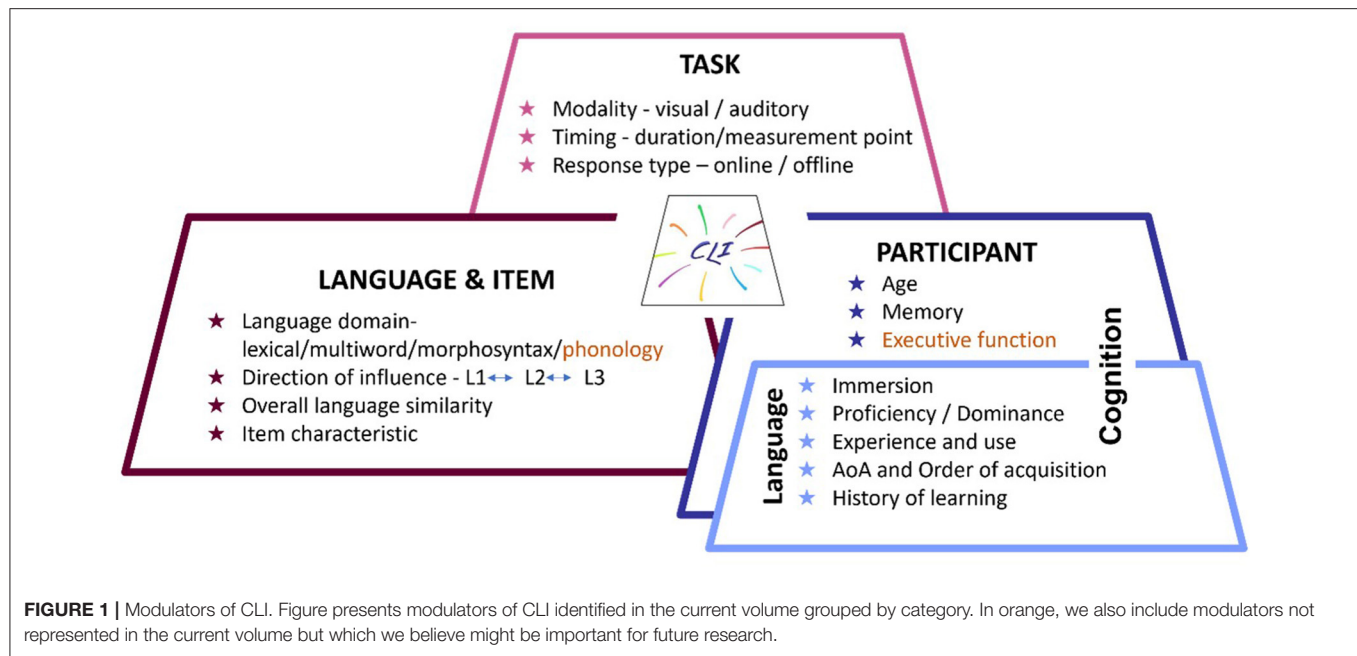
**Received:** 17 March 2022

**Accepted:** 19 April 2022

**Published:** 16 May 2022

### Citation:

Degani T, Prior A and Wodniecka Z  
(2022) Editorial: Modulators of  
Cross-Language Influences in  
Learning and Processing.  
Front. Psychol. 13:898793.  
doi: 10.3389/fpsyg.2022.898793



studies, and whenever both directions of influence were tested, bidirectional effects were reported, although the specific manifestation appears to depend on a range of factors, including duration/measurement point and stage of processing.

## Multi-Word Expressions

Going beyond the single word level, three studies in this volume examined the way in which CLI is modulated in the case of multi word units, which are fixed sequences of words that tend to co-occur within a given language. Of interest, the key question guiding the work presented here was whether knowledge of MWE transfers across languages, and what factors modulate such CLI (Ostwinowska et al.; Du et al.; Pulido). Here too, a variety of methodologies have been implemented; priming of binomials (e.g., knife and fork, Du et al.), eye tracking while reading collocations, and behavioral and ERP measures of collocation processing. Across these studies, there is evidence for CLI in both directions of influence. Thus, L2 processing appears to be affected by CLI from the L1. At the same time, L1 processing is also affected by CLI from the L2, even when speakers are immersed in their L1 environment. Of note, such influence from L2 to L1 was weaker and evident in some measures but not others.

Interestingly, when stimuli with overlapping representations across languages are compared to non-overlapping controls, a congruency effect leads to facilitated processing (Du et al.; Pulido). However, when unique stimuli from the non-target language are artificially translated to the target language, they lead to interference when compared with existing stimuli in the target language (Ostwinowska et al.).

## Morpho-Syntax

Six studies focused on CLI in the domain of morpho-syntax, and examined processing of overlapping and unique syntactic

structures across the languages of bi- and multilinguals. Of these, three studies examined CLI in specific syntactic structures in bilingual children (Kubota et al.; Meir and Janssen) and adults (Vingron et al.). Although the studies differ in the tested modalities and in the specific syntactic structures targeted, all three reach similar conclusions in that how and when CLI is evident in the syntactic domain most likely differs across specific linguistic structures (i.e., Language/Item related modulators). Further, all three studies find that the extent of CLI is modulated by individual variability in speakers' profiles of language use and dominance. Vingron et al. also demonstrate such modulating effects in the timing of CLI. Interestingly, in a study extending the investigation of CLI to trilingual speakers, the extent and timing of CLI was also modulated by participants' profile of language dominance (CLI from L1 vs. L2 in L3 processing; Abbas et al.). Finally, Leon Guerrero et al. and Russak and Zaretsky examined CLI in more ecologically valid setting in school aged children (see also Lipner et al., for intervention with school-aged children). They found that meta-linguistic skill, developed in any of a speakers' languages, can holistically support narrative processing in the target language, be it in comprehension or production.

Across the six studies, proficiency emerged as an important modulator of CLI in the morpho-syntactic domain, as exemplified in the direction of CLI from L1 or from L2 (Abbas et al.; Vingron et al.) and as exemplified in individual differences or change over time (Kubota et al.; Meir and Janssen).

## INTEGRATIVE SUMMARY

Rather than merely documenting the presence of CLI, the current volume explored various factors that might modulate the degree and nature of CLI, namely under what circumstances, for which learners, and in what manner, CLI might be more or less

**TABLE 1** | Overview of studies in the Research Topic.

| Authors               | CLI assessment                                                                                                               | Experimental paradigm                       | Population                                                                                                                                                | CLI facilitation/interference                                                                                                                                                                                                                                                | *Modulators tested (observed)                                                                                                                                                                                       | Direction tested (observed)      |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| <b>Lexical domain</b> |                                                                                                                              |                                             |                                                                                                                                                           |                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                     |                                  |
| Fricke                | Cognate status                                                                                                               | Auditory lexical decision                   | <ul style="list-style-type: none"> <li>English monolingual adults</li> <li>English-Spanish learners (late)</li> <li>Spanish-English (heritage)</li> </ul> | Cognate facilitation                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>Type of background noise (no)</li> <li>Variation in linguistic experience (yes)</li> <li>L1 and L2 proficiency (yes)</li> </ul>                                              | L1 on L2 (yes)                   |
| Hoshino et al.        | Semantic, phonological, or translation word overlap                                                                          | Picture naming in picture-word interference | <ul style="list-style-type: none"> <li>Spanish-English adults in US</li> <li>Japanese-English adults in US</li> </ul>                                     | In both groups: <ul style="list-style-type: none"> <li>Translation facilitation</li> </ul> Only in Spanish-English group: <ul style="list-style-type: none"> <li>Semantic interference</li> <li>Phonological facilitation</li> <li>Phono-translation facilitation</li> </ul> | <ul style="list-style-type: none"> <li>Script (yes)</li> <li>Type of cross-language similarity (yes)</li> </ul>                                                                                                     | L1 on L2 (yes)                   |
| Iniesta et al.        | <ul style="list-style-type: none"> <li>Cognate status</li> <li>Degree of orthographic and phonological similarity</li> </ul> | Writing to dictation                        | English-Spanish adults in US                                                                                                                              | <ul style="list-style-type: none"> <li>Facilitation</li> <li>Interference</li> </ul>                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>Speakers' language background: heritage vs. late learners (yes)</li> <li>Type of cross-linguistic similarity (yes)</li> </ul>                                                | L1 on L2 (yes)<br>L2 on L1 (yes) |
| Lipner et al.         | Lexical depth and breadth as a function of language intervention                                                             | Vocabulary intervention                     | English-Hebrew children in IL                                                                                                                             | Semantic facilitation (knowledge transfer, qualitatively observed)                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>Vocabulary dominance at baseline (yes)</li> <li>Language proficiency (yes)</li> <li>Age of acquisition (yes)</li> <li>Type of language: heritage vs. second (yes)</li> </ul> | L1 on L2 (yes)<br>L2 on L1 (yes) |
| Marian et al.         | Cross-language word similarity                                                                                               | Self-paced paired-associate learning task   | English monolingual adults in US                                                                                                                          | <ul style="list-style-type: none"> <li>Facilitation due to similarity at early stages of learning</li> <li>Interference due to similarity at later stages</li> </ul>                                                                                                         | <ul style="list-style-type: none"> <li>Similarity with native language words (yes)</li> <li>Sequence of learning (yes)</li> </ul>                                                                                   | L1 on L2 (yes)                   |

(Continued)

TABLE 1 | Continued

| Authors                        | CLI assessment                                                                                                                                   | Experimental paradigm                                                                                                        | Population                                                                                                                                                                                               | CLI facilitation/interference                                                                                                                                               | *Modulators tested (observed)                                                                                                   | Direction tested (observed)      |
|--------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| Whitford and Joannis           | Within- and cross-language word form similarity                                                                                                  | Naturalistic paragraph reading, eye-tracking                                                                                 | French-English adults and children in Canada                                                                                                                                                             | Facilitation for similar word forms                                                                                                                                         | <ul style="list-style-type: none"> <li>Age/reading experience (yes)</li> <li>Orthographic neighborhood density (yes)</li> </ul> | L1 to L2 (yes)<br>L2 to L1 (yes) |
| Woumans et al.                 | Cognate status                                                                                                                                   | Written word production (in response to a picture)                                                                           | Dutch-English adults in Netherlands                                                                                                                                                                      | Cognate facilitation                                                                                                                                                        | Type of sentence constraint (yes)                                                                                               | L1 on L2 (yes)                   |
| <b>Multi-word expressions</b>  |                                                                                                                                                  |                                                                                                                              |                                                                                                                                                                                                          |                                                                                                                                                                             |                                                                                                                                 |                                  |
| Du et al.                      | Binomials (congruent/incongruent and translated)                                                                                                 | Primed visual lexical decision                                                                                               | <ul style="list-style-type: none"> <li>Chinese-English adults in New Zealand</li> <li>English-Chinese adults in China</li> <li>English monolinguals in New Zealand</li> </ul>                            | <ul style="list-style-type: none"> <li>Facilitation for congruent collocations</li> <li>No effect for translated collocations</li> </ul>                                    | Language direction (yes)                                                                                                        | L1 on L2 (yes)<br>L2 on L1 (yes) |
| Otwinowska et al.              | L1 collocational calques from L2                                                                                                                 | <ul style="list-style-type: none"> <li>Acceptability judgments</li> <li>Sentence reading, ERPs</li> </ul>                    | Polish-English adults in Poland                                                                                                                                                                          | <ul style="list-style-type: none"> <li>Facilitation for collocational calques in sentence reading and ERP</li> <li>No effect in acceptability judgments</li> </ul>          | <ul style="list-style-type: none"> <li>Type of task (yes)</li> <li>Language proficiency (no)</li> </ul>                         | L2 on L1 (yes)                   |
| Pulido                         | L2 collocations (congruent/incongruent with the L1)                                                                                              | Sentence reading, eye-tracking                                                                                               | English (L1) advanced learners of Spanish (L2) in US                                                                                                                                                     | Facilitation for congruent collocations                                                                                                                                     | <ul style="list-style-type: none"> <li>L2 Chunking ability (yes)</li> <li>Early vs. late reading measures (no)</li> </ul>       | L1 on L2 (yes)                   |
| <b>Morpho-syntactic domain</b> |                                                                                                                                                  |                                                                                                                              |                                                                                                                                                                                                          |                                                                                                                                                                             |                                                                                                                                 |                                  |
| Abbas et al.                   | L3 Grammatical structures non-overlapping with L1, L2 or both                                                                                    | <ul style="list-style-type: none"> <li>English sentence reading, eye tracking;</li> <li>Grammaticality judgements</li> </ul> | Arabic-Hebrew-English trilinguals in Israel                                                                                                                                                              | Interference from L1 and from L2 in L3 processing. No interference when L1 and L2 overlap with each other and differ from L3                                                | <ul style="list-style-type: none"> <li>Language status (yes)</li> <li>Task/measure (yes)</li> </ul>                             | L1 on L3 (yes)<br>L2 on L3 (yes) |
| Kubota et al.                  | <ul style="list-style-type: none"> <li>Genitive structures (partially overlapping)</li> <li>Verb argument structure (non-overlapping)</li> </ul> | Preference / Grammaticality Judgment                                                                                         | <ul style="list-style-type: none"> <li>Japanese-English bilingual children in Japan, immediately after English immersion, and one year later.</li> <li>English monolingual children in the UK</li> </ul> | <ul style="list-style-type: none"> <li>Differences in processing genitive structures (influence by L1)</li> <li>No effect in processing Verb-Argument structures</li> </ul> | <ul style="list-style-type: none"> <li>Proficiency and immersion (yes)</li> <li>Grammatical Structure (yes)</li> </ul>          | L1 on L2 (yes)                   |

(Continued)

**TABLE 1** | Continued

| Authors              | CLI assessment                                                                                                                                                | Experimental paradigm                                                                                                 | Population                                                                                                                                                                                                 | CLI facilitation/interference                                                                                | *Modulators tested (observed)                                                                                                                 | Direction tested (observed)                  |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|
| Leon Guerrero et al. | Linguistic and meta-linguistic skills in L1 and L2                                                                                                            | <ul style="list-style-type: none"> <li>English text reading, eye tracking</li> <li>Comprehension questions</li> </ul> | Spanish-English and monolingual English adolescents in the US                                                                                                                                              | Syntactic integration skill in L1 associated with improved L2 reading and comprehension                      | Text syntactic difficulty (yes)                                                                                                               | L1 on L2 (yes)                               |
| Meir and Janssen     | <ul style="list-style-type: none"> <li>Genitive and Accusative structures in Heritage Language (which do or do not overlap with Societal language)</li> </ul> | Russian oral elicitation task                                                                                         | <ul style="list-style-type: none"> <li>Russian-Hebrew bilingual children in Israel</li> <li>Russian-Dutch bilingual children in the Netherlands</li> <li>Monolingual Russian children in Russia</li> </ul> | Societal language facilitates acquisition of shared Heritage language structures                             | <ul style="list-style-type: none"> <li>Language similarity (yes)</li> <li>Participants' proficiency in the Heritage Language (yes)</li> </ul> | Societal Language on Heritage Language (yes) |
| Russak and Zaretsky  | Linguistic and meta-linguistic skills in L1, L2, and L3                                                                                                       | English oral narrative production                                                                                     | <ul style="list-style-type: none"> <li>Hebrew-English bilingual</li> <li>Arabic-Hebrew-English trilingual children in Israel</li> </ul>                                                                    | Meta-linguistic awareness in Hebrew (L1/L2) associated with improved narrative production in English (L2/L3) | Language group profile (yes)                                                                                                                  | L1/L2 on L3 (yes)                            |
| Vingron et al.       | <ul style="list-style-type: none"> <li>Noun adjective order (partially overlapping)</li> <li>Object-pronouns structure (non-overlapping)</li> </ul>           | English sentence reading, eye tracking                                                                                | <ul style="list-style-type: none"> <li>French-English bilingual adults</li> <li>English-French bilingual adults</li> <li>English monolinguals In Canada</li> </ul>                                         | Faster processing for English violations consistent with French.                                             | <ul style="list-style-type: none"> <li>Grammatical structure (yes)</li> <li>Participants' L2 background and exposure (yes)</li> </ul>         | L1 on L2 (yes)<br>L2 on L1 (yes)             |

\*Some of the effects observed were weak, but we opted to make a binary decision regarding whether modulations and influences were observed. See original papers for more details.

evident. Across 16 independent studies, it becomes clear that there is a high level of interaction between the languages in the multilingual mind, such that the absence of CLI seems to be an exception rather than a norm, although sometimes the observed CLI effects are subtle or weak.

In reviewing the contributions in this volume, we recognize several important common modulators that are linked to the task, the language/item, and the participants of interest. With respect to Task, all studies reviewed here show that whether CLI manifests as facilitation or interference may to a large degree depend on the particular task employed, and especially on what is selected as the baseline against which comparisons are made (see e.g., Du et al. vs. Otwinowska et al.; Hoshino et al.; Iniesta et al.; Woumans et al.). Relatedly, in several cases, dissociations emerged in the patterns of CLI observed as a function of the experimental measure. Manifestations of CLI differed across brain and behavioral indices (Otwinowska et al.), or across offline and more online measures of processing (Abbas et al.).

With respect to Language/Item characteristics, the studies reviewed here convincingly demonstrate CLI across multiple language domains, including lexicon and grammar. However, more work is needed in the domain of phonology as well as in studies including more than a single language domain, to more directly compare the effect of specific modulators across domains. Of note, the studies included in this volume differed in the languages selected as their target (L1/L2/L3), and demonstrate that CLI may take a bidirectional form (Iniesta et al.; Lipner et al.; Whitford and Joanisse; Otwinowska et al.; Vingron et al.). Specifically, dominant L1 influences less dominant L2/L3 processing, but also vice versa, although the latter direction seems weaker (Du et al.). Of note, not all items appear to be affected by CLI to the same extent. This is especially evident in the morpho-syntactic domain, where structure related differences appear to modulate the observed effects (Kubota et al.; Meir and Janssen).

With respect to Participant characteristics, and related to the issue of language direction of CLI noted above, participants' proficiency profile emerges as an important modulator (Abbas et al.; Vingron et al.). Further, language background—namely whether the language was learned as a heritage language or via school-based learning appears to exert an effect (e.g., Fricke). Relatedly, language immersion has been highlighted as crucial for understanding when and how CLI is manifested (Kubota et al.; Meir and Janssen). Indeed, recent work highlights the relevance of patterns of language use as affecting multilingual performance (Gullifer et al., 2018; Beatty-Martínez et al., 2020) and more work is needed to understand their role in possibly modulating

CLI. Additionally, we suggest that future work examine whether individual differences in executive control might also modulate the expression of CLI (Prior et al., 2017).

The body of literature included in this volume highlights the contribution of these three types of modulators, and we therefore advocate for including them in future theoretical and empirical work on CLI. Interestingly, two other issues emerge from the integration of the studies in the current volume. First, evidence for CLI observed at a given point in time might in fact reflect processes that had taken place during learning of the relevant linguistic representations, or may reflect concurrent activation across languages as processing unfolds (for further discussion see Du et al.). Moreover, prior linguistic knowledge may exert differential influence depending on the timing at which it occurs (see Marian et al.). Second, whereas the most common approach to examining CLI focuses on linguistic knowledge/representations, the current volume also documents instances where prior linguistic experience exerted its influence via previously practiced skills (Leon Guerrero et al., Lipner et al., Russak and Zaretsky; for discussion see Hirosh and Degani, 2018).

To conclude, the current volume brings together research from diverse perspectives and domains, once again underscoring the critical role of CLI for understanding multilingual processing. The unique contribution of the current volume is in emphasizing that CLI is not a monolithic phenomenon, but rather varies in systematic ways as a function of task, language and participant. These and related modulators should be embraced in future work on multilingual learning and processing. In particular, we suggest that a concentrated effort examining the effect of a selected modulator (e.g., task demands) across different levels of other modulators (Language/Item and Participant characteristics) may be the most fruitful pathway to move the field forward.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

ZW was supported by the grant SONATA BIS no. 2015/18/E/HS6/00428 from the National Science Centre Poland. AP and TD were supported by grant 340/18 from the Israel Science Foundation.

## REFERENCES

- Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., and Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *J. Exp. Psychol.* 46:1022. doi: 10.1037/xlm0000770
- Cenoz, J., Hufeisen, B., and Jessner, U. (2001). *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, vol. 31. Bristol: Multilingual Matters. doi: 10.21832/9781853595509
- Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., et al. (2018). Bilingual experience and resting-state brain connectivity: impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia* 117, 123–134. doi: 10.1016/j.neuropsychologia.2018.04.037
- Hirosh, Z., and Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: an integrative review. *Psychon. Bull. Rev.* 25, 892–916. doi: 10.3758/s13423-017-1315-7
- Jarvis, S., and Pavlenko, A. (2007). *Crosslinguistic Influence in Language and Cognition*, 1st Edn. London: Routledge. doi: 10.4324/9780203935927
- Prior, A., Degani, T., Awawdy, S., Yassin, R., and Korem, N. (2017). Is susceptibility to cross-language interference domain specific? *Cognition* 165, 10–25. doi: 10.1016/j.cognition.2017.04.006

Puig-Mayenco, E., González Alonso, J., and Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Lang. Res.* 36, 31–64. doi: 10.1177/0267658318809147

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the author and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Degani, Prior and Wodniecka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Cross-Linguistic Influence on L2 Before and After Extreme Reduction in Input: The Case of Japanese Returnee Children

Maki Kubota<sup>1\*</sup>, Caroline Heycock<sup>2</sup>, Antonella Sorace<sup>2</sup> and Jason Rothman<sup>3,4</sup>

<sup>1</sup> Department of Language and Culture, UiT The Arctic University of Norway, Tromsø, Norway, <sup>2</sup> School of Philosophy, Psychology & Language Sciences, The University of Edinburgh, Edinburgh, United Kingdom, <sup>3</sup> Department of Language and Culture, UiT The Arctic University of Norway, Tromsø, Norway, <sup>4</sup> Department of Language and Culture, Nebrija University, Hoyo de Manzanares, Spain

## OPEN ACCESS

### Edited by:

Anat Prior,  
University of Haifa, Israel

### Reviewed by:

Yasuhiro Shirai,  
Case Western Reserve University,  
United States  
Natalia Meir,  
Bar-Ilan University, Israel

### \*Correspondence:

Maki Kubota  
makikubota5@gmail.com

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 May 2020

**Accepted:** 20 August 2020

**Published:** 18 September 2020

### Citation:

Kubota M, Heycock C, Sorace A  
and Rothman J (2020)  
Cross-Linguistic Influence on L2  
Before and After Extreme Reduction  
in Input: The Case of Japanese  
Returnee Children.  
Front. Psychol. 11:560874.  
doi: 10.3389/fpsyg.2020.560874

This study investigates the choice of genitive forms (the woman's book vs. the book of the woman) in the English of Japanese-English bilingual returnees (i.e., children who returned from a second language dominant environment to their first language environment). The specific aim was to examine whether change in language dominance/exposure influences choice of genitive form in the bilingual children; the more general question was the extent to which observed behavior can be explained by cross linguistic influence (CLI). First, we compared the choice of genitive form between monolingual English speakers and bilinguals who had recently returned to Japan from an English speaking environment. Second, we tracked changes in genitive preference within bilingual children, comparing their performances upon return to Japan to those of one year later. Results show that CLI alone is insufficient to explain the difference in genitive evaluation between bilinguals and monolinguals, as well as the intra-group bilingual variation over time. We suggest that both CLI and general processing considerations couple together to influence the changes in genitive preference.

**Keywords:** cross-linguistic influence, genitive (possessive) relations, verb argument construction, Japanese returnees, longitudinal study

## INTRODUCTION

Studies examining (potential) cross-linguistic influence (CLI) provide valuable data for fully understanding how both languages of bilinguals develop and interact over time. In the present study, we define CLI as influence on behavior exhibited by bilinguals that can reasonably be attributed to their other language, that is, not otherwise explainable by developmental processes also noted in monolingual language development (be it delay or acceleration). For example, Japanese learners of English may produce phrases such as *sale's spelling* (rather than *the spelling of "sale"*)—a form that is rarely observed in a developmental stage of English monolingual speakers (Tomiya, 2000). This behavior can reasonably be attributed to CLI from L1 Japanese due to the fact that Japanese only allows pre-nominal possessive construction that linearly maps onto the English s-genitive form.

A large body of literature has been examining the circumstances in which CLI occurs between the two languages of a bilingual. An explanation that figures prominently in this literature is language dominance. On the one hand, studies have shown CLI to take place from the dominant language to the non-dominant language (Yip and Matthews, 2000; Argyri and Sorace, 2007). But others (Müller and Hulk, 2001; Paradis and Navarro, 2003; Sorace and Filiaci, 2006) have instead proposed that the linguistic properties of the two languages—namely, structural overlap and interface condition, discussed further below—determine the occurrence and the directionality of CLI.

The aim of the current study is two-fold. First, in Study 1, we examine the role of linguistic properties in predicting CLI in Japanese-English bilingual children by comparing them to a control group of monolingual peers. By testing the bilinguals within weeks of their return to Japan, we aim to capture their acquired competence in English after significant exposure to native English in an immersion context (mean time of immersion: 4 years). At this point any observable influence from Japanese would be especially significant, speaking to the robustness of CLI effects (i.e., taking place despite ample exposure and high quality of L2 input). Further, the performance of the bilinguals in Study 1 also serves as their own baseline in Study 2, where they are tested after a year of reintegration into Japan. Study 2 thus probes for changes over time within individual speakers, which we also hypothesize will follow a particular pattern induced by CLI effects. A unique aspect of this study is manifested in Study 2, that is, only by studying returnees, can we meaningfully see how CLI and reduction of input interact in a context of L2 attrition. This special context and our longitudinal approach permit us to investigate the effects of a shift in environmental language dominance, as it changes from second language dominant (an English-speaking environment) to first language dominant (Japan).

The studies focus on two grammatical phenomena: the choice of genitive forms and verb/argument order. Starting with the first of these, in English there are two principal ways to express a possessive relationship within a noun phrase: the pre-nominal possessive form, or *s*-genitive (e.g., *the table's leg*), and the post-nominal possessive form, or *of*-genitive (e.g., *the leg of the table*). In Japanese, there is only a pre-nominal genitive, where the pre-nominal possessive is suffixed with the particle *no* (e.g., *Hanako no koppu*; Hanako's cup). In terms of verb/argument order, Japanese is an SOV language whereas English is SVO. These differences lead to predictions about CLI. First, we expected CLI to occur in genitive forms but not verb/argument order (when comparing bilinguals to their monolingual counterparts), due to the fact that genitive forms meet two conditions of CLI (Müller and Hulk, 2001): structural overlap and integration of pragmatic and/or semantic factors, while verb/argument order fulfills neither of these conditions. Second, we expected the effects of CLI from L1 Japanese to L2 English to increase over time after the bilingual returnee children have returned to an L1 dominant environment with minimal L2 exposure.

## Explaining CLI

### Language Dominance

Since many bilingual children are more proficient in or more exposed to one language than the other, some studies have argued that CLI is unidirectional, taking place from the dominant to the non-dominant language (Yip and Matthews, 2000; Paradis, 2001; Argyri and Sorace, 2007; Kupisch, 2007; Nicoladis, 2012). For example, Argyri and Sorace (2007) found that CLI in syntactic structures occurred from English to Greek among bilingual children, but this effect was found only among bilinguals who were dominant in English, not in children who were dominant in Greek. It should be noted that “dominance” is defined in a number of different ways in the literature. For example, Argyri and Sorace (2007) as well as Serratrice et al. (2009) define children's “dominant language” as the majority language of the environment (i.e., Italian in Italy), while Yip and Matthews (2000); Nicoladis (2012) use proficiency measures such as mean length of utterance (MLU) or the Peabody Picture Vocabulary Test (PPVT) to determine the dominant language of bilingual children. It remains the case that no uniform definition exists for this term (for further discussions see Treffers-Daller and Silva-Corvalán, 2016). In this paper we follow the studies that define language dominance in terms of the relative amount of language exposure the child receives in each language.

Tomiya (1999, 2000) is an example of a study that found evidence of L1 CLI due to reduced L2 exposure. In this study, Tomiyama tracked L2 English progress of a Japanese returnee child longitudinally over the course of 33 months. The child was 8 years old at the time of his return to Japan and data was collected once a month using free conversation and a story-telling task. In the second stage of data collection (from 20 months to 33 months), the child used erroneous *s*-genitive forms such as “\*the window's place.” Tomiyama concluded that the inappropriate use of the *s*-genitive is an indicator of L1 CLI, since the genitive form in Japanese resembles the linear order in the *s*-genitive in English. Moreover, studies from Yoshitomi (2007); Snape et al. (2014) reveal that aspectual domain in L2 English showed some signs of attrition after 8–12 months of returning to Japan.

### Linguistic Properties and Processing

An alternative hypothesis for explaining CLI—focusing on the internal structures of the two languages—was first proposed by Hulk and Müller (2000); Müller and Hulk (2001), suggesting that linguistic phenomena subject to CLI must (a) involve two modules of grammar (e.g., syntax/pragmatics) and (b) have similar structures but also be “ambiguous.” Here, ambiguity refers to cases when there is an overlap between the two languages in the sense that one language allows only one form to express a particular function, whereas the other language has two. This also determines the directionality of CLI: the language with one form influences the language with two forms.

Elaborating this idea for adult bilinguals, Sorace and Filiaci (2006) propose that structures that are conditioned by contextual or pragmatic factors are especially difficult to acquire and are also more vulnerable to effects of attrition than structures that only involve syntactic aspects of the language. It is important to note, however, that unlike Müller and Hulk, Sorace and

Filiaci do not make any explicit claims about the directionality or source of CLI, but rather propose that there are different conditions on syntactic realization in bilingual acquisition, which depend to a greater or lesser extent on coordination with “external” (pragmatic, contextual) factors (Sorace, 2011, 2016). The principal empirical test bed for this hypothesis has been the distribution of pronominal forms. For example, in Italian, there are two ways to express pronominal subjects: overt and null pronouns. The choice of these two forms is governed by pragmatic factors—a null pronoun is used when referring to the topic of the previous sentence, whereas an overt pronoun is used to refer to a non-topical antecedent. In contrast to Italian, English only has one form, overt pronouns (e.g., *he*, *she*) to express the same functions. Thus, according to Sorace and Filiaci (2006), English-Italian bilinguals may behave differently from monolinguals in the comprehension and production of pronominal forms in Italian, not only because of the linguistic differences between Italian and English, but also because of the processing load related to linking pronouns to their antecedents in a pragmatically appropriate way in real time.

Sorace (2011, 2016), moving further away from an exclusively generative linguistic framework, elaborated on these ideas by proposing that integration of pragmatic and contextual conditions may be particularly difficult to process for the bilinguals due to the extra cognitive demands it requires. Since cognitive resources are needed to adapt to changing contextual conditions that require one pronoun or the other, bilinguals may experience more competition for cognitive resources, since they also have to inhibit the unwanted language that is not in use. Bilinguals solve this pressure by overextending the scope of the overt pronoun, which is the most explicit form and thus used as a “default” pronominal form to relieve the processing demands caused by the need to integrate pragmatic and/or contextual information. Following this, appealing to cognitive resource allocation as a source of divergence between monolinguals and bilinguals (rather than CLI from a non-null subject language) could better explain why overextension of overt pronouns is found even in bilingual adults (e.g., Lozano, 2006; Margaza and Bel, 2006) as well as bilingual children (Serratrice et al., 2009; Sorace and Serratrice, 2009) speaking two null subject languages. These studies point to the need to examine the interaction between linguistic and non-linguistic (general cognitive) factors in explaining developmental trajectories, particularly for language structures sensitive to ‘external’ contextual conditions.

So far, we have discussed possible explanations for the apparent constrained effects of CLI patterns in bilingual children, focusing on language dominance, linguistic properties, and cognitive load. These factors, however, do not necessarily have to co-occur for CLI to take place. For instance, there is no shortage of empirical evidence showing CLI in the absence of a structural overlap between the two languages of the bilingual and/or for properties that are not clearly interface structures (see e.g., Yip and Matthews, 2000; Nicoladis, 2002; Foroodi-Nejad and Paradis, 2009; White, 2011; Rothman et al., 2019). Nevertheless, the factors of interest here clearly play a role in how CLI obtains, while it is still uncertain how they might interact to drive CLI.

This study contributes, then, by offering a means to tease apart some of these factors. More specifically, it allows us to isolate, under conditions of dramatic reduction in input in one of the two languages, the relative contribution of distinct factors, e.g., overall language dominance versus the types of structures involved, as well as potential interactions.

## Genitive Variation in English

Having now discussed the essential concepts relating to CLI, we now turn to explain the choice of syntactic phenomena that we focussed on in our study, beginning with variation in genitive structures in English. There is considerable debate in the literature as to what determines the choice between the *of*-genitive and *s*-genitive, but also a degree of consensus on some principal factors. These include semantic properties such as animacy and the type of possessive relation, and discourse related factors such as topicality. Animacy is often regarded as the central factor in genitive choice. Several corpus studies (Jucker, 1993; Leech et al., 1994; Gries, 2002; Stefanowitsch, 2003; Rosenbach, 2005) have examined the relative frequency of the two genitive forms when the degree of animacy of the possessor is manipulated. For example, animacy may be treated as a binary category [+/-human] or subcategorized further into “human,” “animal,” “company,” “time,” and “place.” The results of these studies show that animate possessors are more likely to be expressed by the *s*-genitive, while inanimate possessors are more likely to be realized with the *of*-genitive. Similarly, experimental studies such as Rosenbach (2001, 2003) show that the higher the referent is in animacy (e.g., human > animal > object), the more likely it is to occur as an *s*-genitive.

Although extensive research has been conducted on the role of animacy in genitive variation in English monolingual adults, the developmental process of how children acquire this linguistic constraint is still under-explored. One study by Skarabela and Serratrice (2009) investigated whether adults and 4 year-old English monolingual preschool children are aware of the animacy constraint, by using a picture-description syntactic priming task. Their results from the baseline task reveal that both the children and the adults used more *s*-genitive than *of*-genitive to express kinship relationships (e.g., the girl’s mother > the mother of the girl). This suggests that 4-year old children are aware of the animacy constraint in the choice of the two genitive forms. Moreover, this finding accords with Bannard and Matthews (2008)’s conclusion that English-speaking children are aware of the two genitive forms around the age of four.

Other research (Anschutz, 1997; Rosenbach and Vezzosi, 2000; Rosenbach, 2001, 2003; Hinrichs and Szmrecsanyi, 2007) have suggested that the givenness or topicality of the possessor influences the choice of genitive forms. Rosenbach (2001, 2003) demonstrated that [+givenness] and [+definite] referents have a higher likelihood of being expressed using the *s*-genitive. Thus, for example English native speakers are more inclined to use the *s*-genitive for a definite possessor that has been previously mentioned (e.g., *the woman’s body*), and the *of*-genitive for a first mentioned possessor with an indefinite article (e.g., *the body of a woman*).

Another relevant factor is the semantic relationship between the possessor and the possessum. Rosenbach (2001) offers a binary categorisation of the possessive relationship in semantic terms: (a) prototypical relationships which consists of kin terms (e.g., *doctor's son*), body parts (e.g., *girl's hand*) and permanent ownership of concrete things (e.g., *father's car*), and (b) non-prototypical relationships, which cover the remaining possessive cases, including social relations (e.g., *Saint Paul's teacher*), mental/physical states (e.g., *the girl's excitement*) and abstract possession (e.g., *the man's name*) (p. 279). Prototypical relationships have a higher likelihood of being expressed by the *s*-genitive (i.e., [+proto]), while non-prototypical relationships are more likely to be realized by the *of*-genitive (i.e., [−proto]).

A central concern of the literature has been to tease apart the interplay of these factors to determine which have the greatest and which the least influence on the choice of genitive form (Rosenbach, 2014). The framework established by Rosenbach (2001) tested the relative influence of the factors by combining the three factors (animacy, topicality, and possessive relationship) in a hierarchical structure of cells. The summary of the framework is provided in **Figure 1**.

Here, animacy is ranked at the top as the primary factor, followed by topicality, and then the type of possessive relationship. Under this framework, the relative frequency of the *s*-genitive is expected to gradually decrease from the far-left condition [+animate][+topical][+proto] to the far right [−animate][−topical][−proto] and vice-versa for the *of*-genitive. Rosenbach (2001) conducted an empirical study on 56 British native speakers of English to test the validity of this hierarchy. She created a forced-choice task between *of*- and *s*-genitive, controlling for the number of examples for each condition and for the other possible factors that might influence genitive choice. The results confirmed her prediction: there was a steady decrease in the proportion of *s*-genitives from the left to right along the cases defined by this framework.

In a questionnaire-like elicitation task, Rosenbach (2003) counted the frequencies of genitive forms from American and British speakers of English. Their main finding revealed that—similar to the results of the forced choice task by Rosenbach (2001)—there was a steady decrease of both speakers' use of *s*-genitives from [+animate][+topical][+proto] to [−animate][−topical][−proto]. It also found that older American-English speakers used more *s*-genitives than older

British-English speakers, however, there was no significant difference in the relative frequency of *s*-genitives between the younger American-English and British-English speakers. The finding that the use of *s*-genitive with inanimate nouns is more pronounced in American English than British English is supported by several studies (Rosenbach, 2001; Hinrichs and Szmrecsanyi, 2007; Szmrecsanyi and Hinrichs, 2008), however the majority of these studies have examined press language and so how this variation across varieties of English generalizes to the spoken language still remains uncertain.

Although studies on acquisition of English genitive structure in L2 learners are somewhat limited, a corpus-based study by Gries and Wulff (2013) found that learners' genitive choice (between *s*-genitive and *of*-genitive) are moderated by their L1. While Chinese speakers used English genitive forms similarly to their English monolingual counterparts, German speakers showed stronger bias toward the use of *s*-genitives. However, Ghilzai (2014) using a speeded judgment task found that German speakers used less *s*-genitives than the monolingual controls, specifically in the [+animate][+topical][+proto], [+animate][+topical][−proto] [+animate][−topical][+proto] conditions as described in the **Figure 1** above. Such contradictory results, however, may be an artifact of the methodologies used in these studies—Gries and Wulff (2013) used a corpus and examined linguistic production while Ghilzai (2014) investigated interpretation/comprehension through eliciting judgements.

As mentioned above, it is important to note here that interpretation and processing of English genitives require integration of multiple factors including pragmatic (topicality) and semantic (animacy and prototypicality) information, which are hypothesized to be variable and open to the effects of CLI to different extents (Sorace, 2011, 2016).

## Genitive Structure in Japanese

While English has two genitive constructions, with the choice between them influenced by various factors as discussed in the previous subsection, Japanese has only one construction: the pre-nominal *no* construction. The genitive case marker *no* stands between the possessor and the possessum (e.g., “Hanako *no* penn” “Hanako's pen”) and thus the construction has a similar linear order to the *s*-genitive in English. There have been argued to be more than fifteen types of semantic relationship that can hold between the two noun phrases in the Japanese genitive

| [+animate]          |                          |                    |                       | [−animate]             |                         |                       |                    |
|---------------------|--------------------------|--------------------|-----------------------|------------------------|-------------------------|-----------------------|--------------------|
| [+topical]          |                          | [−topical]         |                       | [+topical]             |                         | [−topical]            |                    |
| [+proto]            | [−proto]                 | [+proto]           | [−proto]              | [+proto]               | [−proto]                | [+proto]              | [−proto]           |
| the boy's eyes      | the mother's future      | A girl's face      | a woman's shadow      | the chair's frame      | the bag's contents      | a lorry's wheels      | a car's fumes      |
| the eyes of the boy | the future of the mother | the face of a girl | the shadow of a woman | the frame of the chair | the contents of the bag | the wheels of a lorry | the fumes of a car |

**FIGURE 1** | Genitive framework from Rosenbach (2001).

construction (Teramura, 1991). Importantly, the Japanese *no*-genitive maps straightforward onto the *s*-genitive in English and both are cliticized morphological exponents right-attached to the possessor noun phrase.

Japanese children start producing the *no*-genitive at an early age (2;2–2;4), and in fact *no* is one of the earliest case particles that they acquire (Clancy, 1985). According to the systematic review of acquisition order of grammatical morphemes in Luk and Shirai (2009), Japanese learners of English acquire the *s*-genitive construction at an earlier stage than other grammatical morphemes such as articles, past-tense morpheme in regular verbs, and third person singular *-s*. This finding has been obtained in studies with Japanese-English children (Hakuta, 1976) as well as adults (Nuibe, 1986; Shirahata, 1988; Izumi and Isahara, 2004). The authors conclude that the linear similarity between the English *s*-genitive and Japanese *no*-genitive allows for positive L1 transfer to occur from Japanese to English.

## Typological Differences Between Japanese and English

Several typological differences exist between Japanese and English. Most importantly, Japanese is an SOV language while English is SVO. Relatedly, English is a head + modifier language in which extensive expansion generally occurs to the right of the non-expandable element, while Japanese is a modifier + head language, where extensive expansion occurs to the left of the non-expandable element. Additionally, English is a mildly synthetic language, while Japanese is analytic in the sense that it has no noun inflection but has a complex system of verb inflection.

## STUDY 1

Our study first examines bilinguals' (Japanese-English) and monolinguals' (English) knowledge of English genitive constructions. In addition, we also examine their knowledge of the word order between the verb and its arguments (subject, object, and indirect object). The choice of the order of the verb and its arguments, specifically in the contexts we use, makes an ideal comparison to the genitive form, as the verb/argument order sentences used in our study lack structural overlap between English (SVO) and Japanese (SOV) and the choice between them is relatively insensitive to interface conditions of an external nature. Thus, the prediction, under all accounts, is that this aspect of word order will not be (easily) affected and thus these conditions serve as a controlled baseline. In Study 1, we first look at preferences for genitive structure (*s*-genitive versus *of*-genitive) and for verb/argument orders, comparing the Japanese-English bilinguals to English monolinguals to see whether there are any differences in their evaluation of these constructions, and if so, in what contexts. The research questions for Study 1 are as follows:

- (1) Are there any differences in the evaluation of genitive forms and verb/argument order between bilingual and monolingual children?
- (2) If so, does an account of CLI based on linguistic properties suffice to explain observed differences?

First, we do not expect monolinguals and bilinguals to behave differently in their responses for verb/argument order. This is because these aspects of verb/argument order in Japanese (SOV) and English (SVO) do not exhibit structural overlap (Müller and Hulk, 2001) and are largely unaffected by non-syntactic factors (Sorace and Filiaci, 2006). We would expect that the default verb/argument order for both languages was acquired by the onset of testing. However, for the genitive items, as described previously, the children need to know that (a) English has two forms to express possession while Japanese only has one and (b) there are multiple non-syntactic (pragmatic and semantic) factors that influence the choice of the two English genitive forms. The genitive conditions are thus hypothesized to be vulnerable to CLI<sup>1</sup> creating a context in which non-native like outcomes for bilinguals are expected. Thus, we hypothesize that the bilinguals will behave differently from the monolinguals in their evaluation of genitives.

In order to assess their preference for genitive forms, we used the framework established by Rosenbach (2001) as discussed earlier. To make the experiment manageable for children, we used four out of the eight conditions in the framework. Specifically, we restricted ourselves to the two conditions on the far left and the two conditions on the far right of the genitive framework in **Figure 1**: [+animate] [+topical] [+proto], [+animate] [+topical] [−proto], [−animate] [−topical] [+proto], [−animate] [−topical] [−proto]. The test conditions and items in our study are discussed in detail in the methodology section. We predict that overextension of *s*-genitives will be manifested in all conditions, not least because if Japanese were to exercise some influence, the *s*-genitive is the only form that overlaps structurally with the corresponding Japanese construction, at least at the surface, given its linear order.

## Methodology

### Participants

#### Bilingual group

The bilingual group consisted of 36 Japanese-English bilingual children (21 female; 15 male), who acquired English as a second language in a native-English speaking environment outside Japan. The average age of the bilinguals was 9;8 (range 7;6–13;0, *SD* = 1.42). All of the bilingual participants had very minimal exposure to English before leaving Japan. All of the bilingual children's parents speak Japanese as their native language and the children were exposed to Japanese from birth. Thus, the age of onset of L2 acquisition was the point at which the bilinguals moved to the foreign environment: the average was 5;0 (range 1;0–9;6, *SD* = 2.5). The average length of residence in the foreign country was 4 years (range 2;0–9;9, *SD* = 2.0). Unlike typical Japanese children, the participants learned English through living in a foreign country and attending schools with English as a medium of instruction. Seventeen participants spent their time away from Japan in a country where English is the

<sup>1</sup>We should note here that the Interface Hypothesis has evolved away from a simple distinction between "internal" and "external" interfaces (see Sorace, 2016), toward a recognition that a range of different conditions can influence grammatical realization, some of which may be more taxing than others from a processing point of view.

majority language (United States:  $N = 11$ , United Kingdom:  $N = 5$ , Canada:  $N = 1$ ), and the other 19 participants attended international schools in countries where English is not the official language (Netherlands:  $N = 1$ , France:  $N = 2$ , Singapore:  $N = 5$ , Thailand:  $N = 1$ , France:  $N = 2$ , Israel:  $N = 1$ , Malaysia:  $N = 2$ , Vietnam:  $N = 1$ , Indonesia:  $N = 1$ , China:  $N = 2$ , Poland:  $N = 1$ ), but still received education in English. Although the latter group was exposed to a third language other than Japanese and English, none of the parents reported that their children could actually hold a conversation in the third language. The bilingual participants were recruited from an English maintenance course offered from JOES (Japan Overseas Educational Services). We administered the Bilingual Language Experience Calculator (BiLEC; Unsworth, 2016) to elicit language background information about the bilinguals, which will be further discussed in the context of Study 2 when we examine the change in bilinguals' syntactic preferences over time.

### Monolingual group

The monolingual group consisted of 35 children (Mean age = 9;4, range 7;0–13;9,  $SD = 1.6$ , 15 female). The monolingual children spoke English as their L1 and had very minimal exposure to any L2 (only in language classes at school once a week). The monolingual group was matched to the bilingual group in terms of age and socio-economic status (SES), which was measured by the mother's final education. All the mothers of the children who participated in this research were educated to Bachelor level or higher. The monolinguals were recruited in Edinburgh, United Kingdom and the majority of them were exposed to British English.

### Instruments

An untimed, binary forced-choice task was developed by the researchers. The task consisted of 16 genitive items and 16 verb/argument order items. We describe the genitive items first. As mentioned earlier, four out of the eight possible combinations of factors defined in Rosenbach's (2001) framework were used as conditions in this study: [+animate][+topical][+proto], [+animate][+topical][−proto], [+animate][−topical][+proto], [+animate][−topical][−proto]. We only used four conditions since having four items for each of eight conditions (32 genitive items in total) would have resulted in a long task that would have been too demanding for the children.

Examples of each condition are presented in **Figure 2** (see **Supplementary Data Sheet 1: Table 1** for all items). In the [+animate] [+topical][+proto] example, the possessor is animate (i.e., *girl*) and topical (mentioned in the previous sentence). Further, the relationship between the possessor and the possessum (i.e., *hand*) is prototypical as it expresses the inalienable possession of body parts. In the [+animate][+topical][−proto] example, the possessor is animate and also topical, but the relationship is non-prototypical as the possessum is an abstract object (i.e., a name). The same logic applies for the other two conditions. For convenience, we will label the [+animate][+topical][+proto] as “strong *s*-genitive,” [+animate][+topical][−proto] as “weak *s*-genitive,” [−animate][−topical][+proto] as “weak *of*-genitive,”

and [−animate][−topical][−proto] as “strong *of*-genitive” conditions. In the “strong” conditions, all factors favor the same form, whether *s*-genitive or *of*-genitive; in the “weak” conditions the prototypicality factor has the value that influences in the opposite direction to the other two (animacy and topicality).

The verb/argument order items were grouped into three conditions<sup>2</sup>, including paired structures with one grammatical and one ungrammatical order. The ungrammatical sentences were created by manipulating the position of the subject, verb, object and (where present) indirect object (O: object; DO: direct object; IO: indirect object) as illustrated in the examples in **Figure 3** (see **Supplementary Data Sheet 1: Table 2** for all items). These verb/argument orders (with the exception of SVDOIO) are grammatical in Japanese.

### Procedure

Two puppets, a male and a female, were presented on a PowerPoint screen. Each puppet read the target sentence using either the *of*- or *s*-genitive structure for the genitive items. For example, the female puppet would say: “*A room's darkness can make little children scared*”, whereas the male puppet would say: “*The darkness of a room can make little children scared*.” The same procedure was taken for the verb/argument order items. The sentences spoken by the male puppet were recorded by a male native speaker of American English, whereas the female puppet was voiced by a female native speaker of British English. We used speakers of different dialects since some children in the bilingual group were educated through the British system, while others attended schools with an American educational system.

All participants were seen individually by the researcher in a quiet room, either at home or at school. They were placed in front of a computer screen with the PowerPoint presentation as in **Figure 4**. They were then asked to listen to the pre-recorded instructions and have one practice trial. During the practice trial, they were asked to choose the puppet that spoke better English. The children were reminded to not base their decisions on phonological factors such as accent or pronunciation. In the practice trials, they were asked to explain their decisions, and if the children's explanations were related to phonological factors, they were reminded again to focus on what the puppet actually said, and not on how he/she said it. They were also allowed to hear the sentences again if they wished to, but not more than twice. Following the practice trial, 32 trials (16 genitive and 16 verb/argument order) were presented in random order. All of the responses were recorded on paper by the investigator. The position of the puppets (i.e., left or right of the screen), the puppet that started speaking first, and the amount

<sup>2</sup>In the initial study, we tested for an additional OSV condition such as *I like apples but oranges I hate* (OSV) vs. *I like apples but I hate oranges* (SOV). We have excluded this condition from the analysis since OSV verb/argument order in English is grammatical to the extent that it can be interpreted as involving an operation of “fronting/topicalization” that is felicitous only in a restricted range of discourse contexts (see e.g., Pullum, 1977). That is, we cannot be certain that the interpretation of OSV verb/argument order in our experiment does not involve more than one module of grammar (i.e., syntax), while the other three conditions in **Figure 3** are clearly erroneous sentences and evaluating them does not require integration of semantic or pragmatic features. Eliminating this one condition from the analysis did not change the significance of the results in either Study 1 or Study 2.

| [+animate]                                                                                         |                                                                                                            | [-animate]                                                                               |                                                                                                    |
|----------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| [+topical]                                                                                         |                                                                                                            | [-topical]                                                                               |                                                                                                    |
| [+proto]                                                                                           | [-proto]                                                                                                   | [+proto]                                                                                 | [-proto]                                                                                           |
| <b>strong <i>s</i>-genitive</b>                                                                    | <b>weak <i>s</i>-genitive</b>                                                                              | <b>weak <i>of</i>-genitive</b>                                                           | <b>strong <i>of</i>-genitive</b>                                                                   |
| <i>s</i> -genitive:<br>A bee stung a girl. <u>The girl's hand</u> was swollen for many days.       | <i>s</i> -genitive:<br>A new teacher came to our school. But nobody knew <u>the teacher's name</u> .       | <i>s</i> -genitive:<br>I banged my toe on <u>a table's leg</u> and it hurts a lot.       | <i>s</i> -genitive:<br>It's annoying when people start laughing before <u>a joke's end</u> .       |
| <i>of</i> -genitive:<br>A bee stung a girl. <u>The hand of the girl</u> was swollen for many days. | <i>of</i> -genitive:<br>A new teacher came to our school. But nobody knew <u>the name of the teacher</u> . | <i>of</i> -genitive:<br>I banged my toe on <u>the leg of a table</u> and it hurts a lot. | <i>of</i> -genitive:<br>It's annoying when people start laughing before <u>the end of a joke</u> . |

FIGURE 2 | Examples of genitive items.

of *of*- and *s*-genitive sentences spoken by each puppet were all counterbalanced.

## Analysis

In order to examine whether there are differences in the choice of genitives and verb/argument order between monolinguals and bilinguals, we constructed two models using Generalised Linear Mixed Effect Model (GLMER) with logit link. Both models included binary response as a dependent variable and Group (Bilingual, Monolingual) and Condition (genitive model: strong *s*-genitive, weak *s*-genitive, strong *of*-genitive, weak *of*-genitive; verb/argument order model: SOV, S-IO-DO-V, S-V-DO-IO) as predictors. For the genitive model, *s*-genitive responses were coded as 1 and *of*-genitive as 0. For the verb/argument order model, SVO and S-V-DO-IO were coded as 1 and others as 0. We included Subject and Item as random intercept (adding by-Subject and by-Item random slope for Condition did not improve the overall fit of the model). For the genitive model, the reference level was set to “bilinguals” for the Group and “strong *of*-genitive” for the Condition variables. For the verb/argument order model,

the reference level was set to “bilinguals” for the Group and “SOV” for the Condition variables.

## Results

### Genitive form

The estimated coefficients of the genitive model are presented in the **Supplementary Data Sheet 2: Table 1**. As shown clearly in the comparison of mean percentages of *s*-genitive choice between monolinguals and bilinguals from the first round of data collection (**Figure 5**), significant differences in the evaluation of genitive forms were found for two conditions: the weak *of*-genitive condition and the weak *s*-genitive condition. Pairwise comparison (Tukey's test) demonstrates that the bilinguals used more *s*-genitives in the weak *of*-genitive condition than the monolinguals ( $E = 0.88$ ,  $SE = 0.24$ ,  $z = 3.56$ , adjusted  $p = 0.008$ ). In contrast, the monolinguals' preference for *s*-genitive was higher than the bilinguals in the weak *s*-genitive condition ( $E = -0.99$ ,  $SE = 0.27$ ,  $z = -3.58$ , adjusted  $p = 0.008$ ).

### Verb/argument order

The estimated coefficients of the verb/argument order model are presented in the **Supplementary Data Sheet 2: Table 2**. There are no significant differences between monolinguals and bilinguals in their evaluation of verb/argument order, given the lack of significant interactions between Group and Condition ( $p$ 's  $> 0.79$ ). Moreover, as shown in **Figure 6**, the performance on all verb/argument order conditions was at near-ceiling (SVDOIO) or at ceiling (SOV, SIODOV) for both monolinguals and bilinguals.

## Discussion

In Study 1, we predicted that bilinguals would behave differently from monolinguals in their evaluation of the two types of genitives, but not in their evaluation of the

**TABLE 1** | Summary of BiLEC variables split by language and time; “Abroad” indicates percentage of language exposure when the children lived in the English majority language environment and “Japan” indicates percentage of exposure upon returning to Japan.

|      | L2 English |       | L1 Japanese |       |
|------|------------|-------|-------------|-------|
|      | Abroad     | Japan | Abroad      | Japan |
| Mean | 46.8       | 4.5   | 53.2        | 95.5  |
| SD   | 12.1       | 3.2   | 10.8        | 8.5   |
| Min  | 26.5       | 0     | 17.5        | 28.0  |
| Max  | 82.4       | 20.5  | 61.0        | 92.4  |

| SOV                                         | SIODOV                                          | SVDOIO                                                |
|---------------------------------------------|-------------------------------------------------|-------------------------------------------------------|
| SVO:<br>My dad reads newspapers every day.  | SVIODO:<br>My uncle bought me a bag last week.  | SVIODO:<br>Our teacher taught us the alphabet today.  |
| *SOV:<br>My dad newspapers every day reads. | *SIODOV:<br>My uncle me a bag last week bought. | *SVDOIO:<br>Our teacher taught the alphabet us today. |

**FIGURE 3** | Examples of verb/argument order items.

different verb/argument orders. More specifically, it was expected that bilinguals would over-extend *s*-genitives compared to their monolingual counterparts. As predicted, bilinguals and monolinguals did indeed behave similarly on their evaluation of verb/argument order; however, our results do not bear out the predictions pertaining to the genitive entirely. Recall that the bilinguals not only preferred to use *more s*-genitive in the weak *of*-genitive condition, in line with our predictions, but they also chose *less s*-genitive in the weak *s*-genitive condition when compared to the monolinguals. That is, bilinguals behaved differently from monolinguals in the two “weak” conditions—those for which monolinguals weakly favor either the *s*-genitive or the *of*-genitive, respectively—but did not differ in the “strong” conditions.

The fact that the monolinguals were predominantly exposed to British English in the United Kingdom, while the bilingual group were exposed to different varieties of English dialects may have had some effects on the observed differences in genitive preference between monolinguals and bilinguals. Given that the use of *s*-genitive with inanimate nouns has been found to be more common in American English than British English (Rosenbach, 2001, 2002, 2003; Hinrichs and Szmrecsanyi, 2007; Szmrecsanyi and Hinrichs, 2008), one could expect monolinguals (who were mainly exposed to British English) to choose *less s*-genitives with inanimate possessor (i.e., weak *of*-genitive and strong *of*-genitive conditions) than the bilinguals, in which at least one-third of them were predominantly exposed to American-English. Our results show that indeed, bilinguals used *more s*-genitive in the weak *of*-genitive condition than the monolinguals but behaved similarly in the strong *of*-genitive condition. Moreover,

dialectal differences cannot account for the differences in the weak *of*-genitive condition, as the bilinguals preferred *less s*-genitive in the weak *s*-genitive condition. The limitation of this study is that the monolingual group was homogenous in terms of the English dialect they were exposed to, which makes it difficult to tease apart the influence of dialectal differences vs. CLI from the native language. Future studies should either keep the English dialect consistent between two groups or include speakers from various English dialects for both groups when comparing the genitive form use/interpretation between monolinguals and bilinguals.

These results suggest that bilinguals were able to choose the “preferred” genitive form in the “strong” conditions, regardless of whether it is a context that induces a strong preference for the *s*-genitive or for the *of*-genitive. That is, in the conditions that have consistent [+factors] or [−factors] (i.e., strong *of*-genitive: [−animate][−topical][−proto] or strong *s*-genitive: [+animate][+topical][+proto]), the bilinguals are not different from the monolinguals in their choice of genitives. That is, in these contexts—the polar ends of Rosenbach’s continuum—the relevant feature configurations provide the strongest cluster sets biasing one or the other form. Recall that in all conditions both the *s*-genitive and *of*-genitive are grammatical. What is at stake is their relative likelihood of being chosen. Since bilinguals perform, in general, closer to monolinguals when optionality is reduced, this could account for the data we obtain. In other words, if there is a continuum of optionality (stronger versus weaker conditions in Rosenbach’s terms), we should note less bilingual divergence where things appear more categorical as is the case when the relevant set of feature values is at one or the other end of the continuum. However, when one factor (prototypicality) is in conflict with the other factors (topicality and animacy) as in the weak conditions (weak *of*-genitive: [−animate][−topical][+proto] or weak *s*-genitive: [+animate][+topical][−proto]), the bilinguals appear to have more difficulties in selecting the preferred structure in line with the monolinguals as the evaluation of the feature set is more complex and requires more processing resources. Taken together, the findings do not indicate an overextension of *s*-genitives *per se*, but rather demonstrate that the bilinguals behave differently from monolinguals in conditions that require processing of conflicting semantic and pragmatic factors. Consequently, this suggests that the differences in the choices of highly proficient bilinguals (recall they were tested soon after on average 4 years of

**TABLE 2** | Summary of verbal fluency performance split by language and time; “Round 1” indicates children’s performance at first round of testing and “Round 2” indicates their performance at second round of testing.

|      | L2 English |         | L1 Japanese |         | Relative Proficiency |         |
|------|------------|---------|-------------|---------|----------------------|---------|
|      | Round 1    | Round 2 | Round 1     | Round 2 | Round 1              | Round 2 |
| Mean | 13.60      | 15.6    | 15.40       | 17.71   | 1.81                 | 2.38    |
| SD   | 4.22       | 3.36    | 4.99        | 5.32    | 4.89                 | 5.39    |
| Min  | 8          | 7       | 8           | 10      | 8                    | 15      |
| Max  | 27         | 22      | 34          | 31      | −12                  | −8      |

Relative proficiency is the difference between Japanese and English performance.

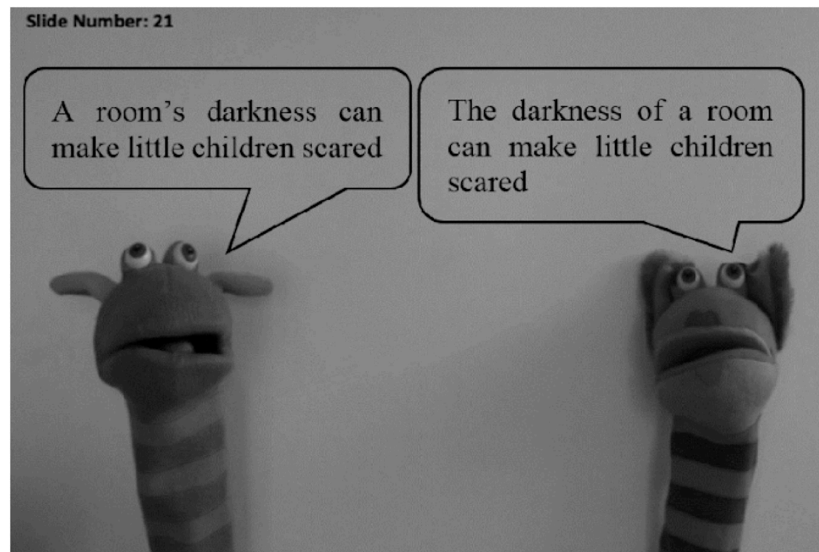


**FIGURE 4 |** PowerPoint presentation of the forced-choice task.

immersion in a native English environment) and monolinguals cannot solely be attributed to CLI due to the internal structure of the two languages, but, in this case, also depends on the relative complexity of semantic and pragmatic integration, consistently with the Interface Hypothesis (Sorace, 2016).

## STUDY 2

Study 2 examines the same bilingual children a second time, comparing Round 2, one year after the first testing, to their own performance from Round 1 (on average within the first weeks of returning to Japan). The results are compared to determine whether there is any change in the choice of genitive forms and verb/argument order, using each bilingual individual as their own baseline. The research questions for Study 2 are as follows:

- (3) Are there any changes in the bilingual returnee children's evaluation of genitive forms and verb/argument orders over time?
- (4) If so, can the observed change be explained by CLI from the dominant to the non-dominant language?

If CLI in the genitive structure is (partially) due to language dominance, then the prediction is that there should be unidirectional CLI from L1 Japanese to L2 English in bilingual returnee children that will increase over time after return to the Japanese environment. This would be due to their L1 Japanese becoming increasingly more dominant after their return to the L1 environment. Since the Japanese *no*-genitive resembles the linear order of *s*-genitive in English and both are cliticized morphological exponents right-attached to the possessor noun phrase, we expect the preference for *s*-genitives to increase across the four conditions. In terms of verb/argument order, there are two possible predictions. If language dominance

alone is “enough” for CLI to occur irrespective of the nature of the underlying structural representation, then we expect children to also accept more erroneous verb/argument order (e.g., SOV) in English over time. However, if some degree of (ambiguous) structural overlap of linguistic properties across the two languages and language dominance are necessary for CLI to take place (Müller and Hulk, 2001) or if basic word order is particularly resilient, then it would be expected that change in language dominance would only affect genitive forms, but not verb/argument order.

## Methodology

### Participants

#### *Bilinguals*

Of the original group of 36, two participants' data were not recorded due to technical issues, leaving 34 in the second round of testing. The average time that elapsed between Round 1 testing and Round 2 was 12 months (range 10–13 months,  $SD = 0.64$ ). Recall that in the first test session, the average age of the bilinguals was 9;8 (range 7;6–13;0 years,  $SD = 1.42$ ). The average age at time of testing in Round 2 is 10;8 (range 8;6–14;0 years,  $SD = 1.42$ ), reflecting a true year between testing sessions.

As mentioned in the Methodology section in Study 1, the Bilingual Language Experience Calculator (BiLEC; Unsworth, 2016) was administered to the parents twice in order to elicit information about quantitative language exposure of the bilingual children in each language and history of language use. The first administering was done at the time of the first testing session and focused on exposure and use of Japanese and English during the stay abroad. The second focused on the distribution of Japanese and English since the time of return. Bringing the two together, it is possible to measure the change at the individual and group level of exposure and engagement with both languages

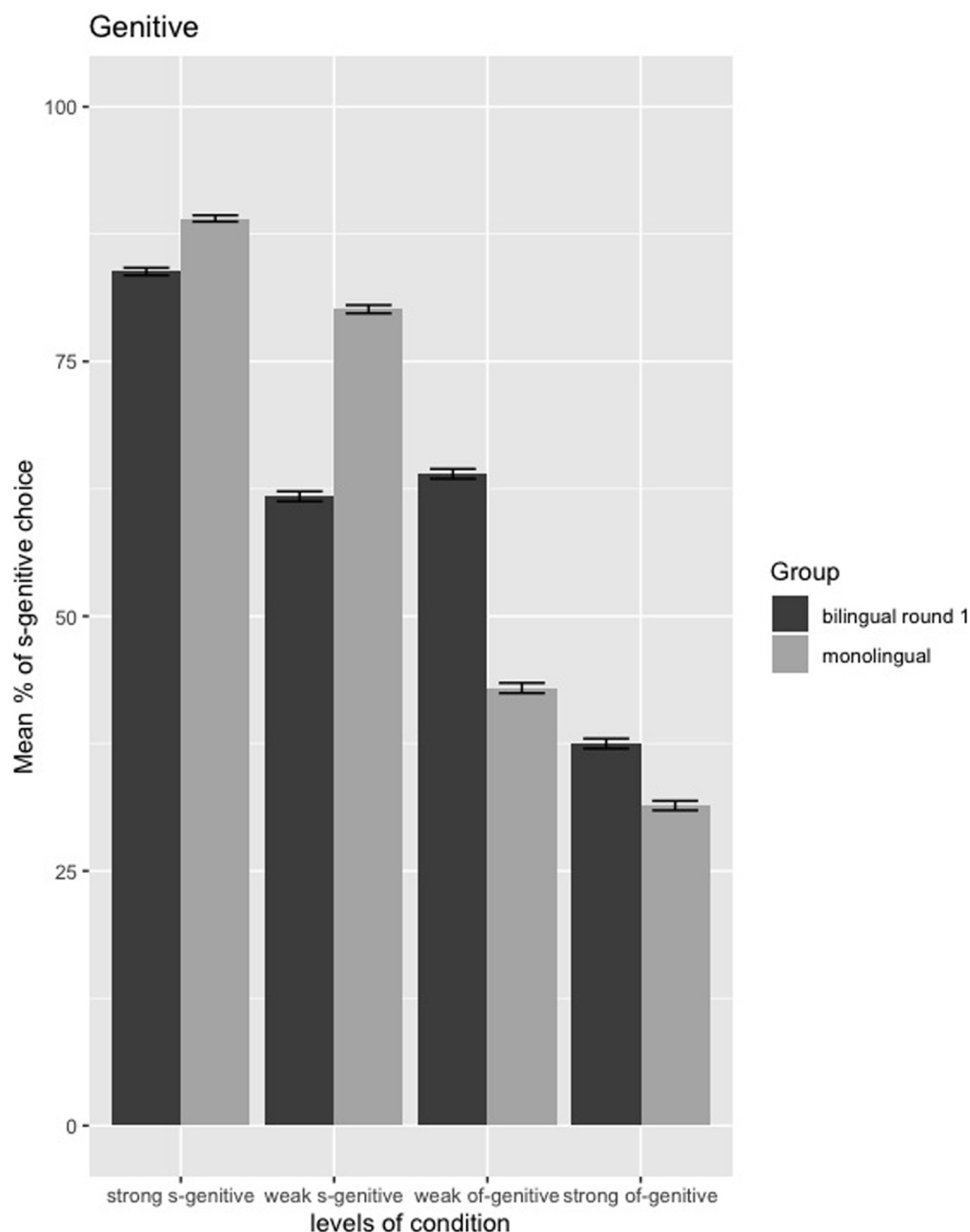


**FIGURE 5 |** Mean percentages of s-genitive choice in the genitive forced-choice task between monolinguals and bilinguals from first round of testing; error bars = standard error.

upon return to Japan. The relevant information is summarized in **Table 1** above.

As shown in **Table 1**, the children's average English exposure decreased from 46.8% while abroad to 4.5% upon return to Japan, a drop of 42.3%. Their Japanese exposure at home and at school increased in an inverse proportion (from 53.2% to 95.5%). It is important to point out here, the bilingual children were not "English-dominant" even when living abroad, since they received around half of their exposure in English and the other half in Japanese. Rather they

were receiving balanced L1 and L2 exposure. However, when returning to Japan, we can see that the exposure they received was clearly Japanese-dominant, with having 95.5% of their exposure in Japanese.

### Instruments and Procedure

We used the same forced-choice task in Study 1 to examine the changes over time in the bilingual children's preferences for genitive and verb/argument order. The only difference in the task was that lexical changes were made for each

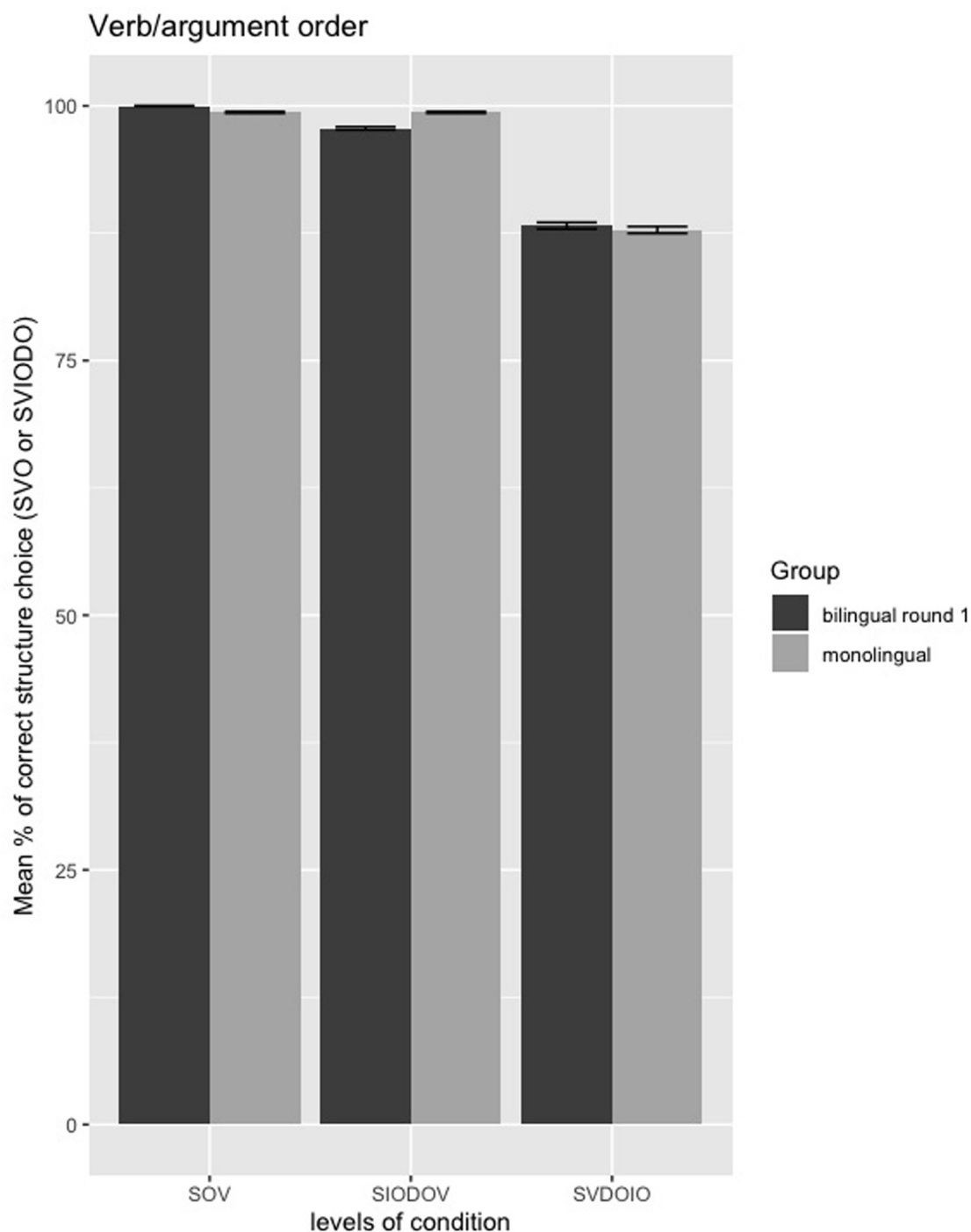


**FIGURE 6 |** Mean percentages of correct/preferred structure choice (SVO or SVIODO) in the verb/argument order forced-choice task between monolinguals and bilinguals from first round of testing; error bars = standard error.

item in order to reduce learning effects in the second round of testing. For example, “*The darkness of a room/a room’s darkness can make little children scared*” was changed to “*The darkness of a room/a room’s darkness can make people anxious.*” The target phrase (e.g., *the darkness of a room/the*

*room’s darkness*) remained the same in the first and the second test. In the second test, a female native speaker of Canadian English voiced the female puppet and a male native speaker of British English the male puppet, since some children were educated through the British system and others

American. The procedure was exactly the same as the first round of testing.

In order to measure the bilingual children's relative proficiency, we administered a verbal fluency task at both first and second round of testing for the bilingual children. The participants were asked to name either (1) animals or (2) fruits and vegetables in English or Japanese. Half of the bilinguals named animals in English and fruits and vegetables in Japanese, and vice-versa for the other half of bilingual participants. For all participants, a timer was set to 1 min by the researcher and the participants were all given approximately ten seconds after they listened to the instruction to start the task. Their responses were recorded on a voice-recorder and later transcribed by two research assistants. The total number of unique words was calculated for each participant, and the difference between Japanese and English scores for each participant was used as a measure for their relative proficiency. The higher the values are, the more proficient they are in Japanese and vice versa for English.

## Analysis

We constructed two GLMER models: one with genitive responses and the other with verb/argument order responses. In the genitive model, we included binary response as a dependent variable and Time (Bilingual Round 1, Bilingual Round 2), Condition (genitive model: strong *s*-genitive, weak *s*-genitive, strong *of*-genitive, weak *of*-genitive), Age of L2 onset (AoO)<sup>3</sup>, L2 exposure difference (i.e., the difference in L2 exposure when they lived in an L2 majority language environment vs. back in the L1 environment), and Relative proficiency (at first round of testing) as predictors. Each language background variables (i.e., AoO, L2 exposure difference, Relative proficiency) were included as a three-way interaction between Time and Condition, in order to examine whether the change in genitive preference for each condition is affected by bilinguals' language experience. We did not include background variables as predictors for the verb/argument model, since we did not expect these variables to influence the change of preference in verb/argument order. In addition, because the bilingual children performed at ceiling for first and second round of testing on the verb/argument items, including a three-way interaction among Condition, Round, and the three background variables resulted in model conversion errors.

For the genitive model, *s*-genitive responses were coded as 1 and *of*-genitive as 0. For the verb/argument order model, SVO and S-V-IO-DO were coded as 1 and others as 0. For the genitive model, the reference level was set to "Bilingual Round 1" for the Time and "strong *of*-genitive" for the Condition variables. For the verb/argument order model, the reference level was set

to "Bilingual Round 1" for the Time and "SOV" for the Condition variables.

## Results

### Relative proficiency

The results of their verbal fluency performance at first and second round of testing is presented in **Table 2**.

Children performed better in Japanese than English for both first and second round of testing, and they increased their scores from first to second round of testing for both English and Japanese. They also appear to have relatively balanced L1 and L2 proficiency, although their dominance in Japanese proficiency increases from first to second round of testing. A categorical classification of relative proficiency is determined through calculating whether there is a difference greater than 1 standard deviation between Japanese and English scores (Unsworth, 2003). Following this categorization, 4 children were English-dominant, 13 children were Japanese-dominant, and 19 children were balanced bilinguals in the first round of testing. In the second round of testing, only 1 child was English-dominant, 10 children were Japanese-dominant, and 25 children were categorized as balanced bilinguals.

### Genitive form

The model summary (see **Supplementary Data Sheet 2: Table 3**) shows that there is an interaction between Group and Round ( $E = 2.50$ ,  $SE = 0.54$ ,  $z = 4.62$ ,  $p \leq 0.001$ ), and the pairwise comparison (Tukey's test) reveals that there was a difference in the genitive evaluation from first to second round of testing in the weak *s*-genitive condition only ( $E = -1.85$ ,  $SE = 0.34$ ,  $z = -5.34$ , adjusted  $p \leq 0.001$ ). The bilinguals showed a greater use of the *s*-genitive in the weak *s*-genitive condition after a year spent back in the Japanese environment, as also illustrated in **Figure 7**.

There were no significant two-way interactions between Round and Background variables (AoO, L2 exposure, and Relative proficiency) ( $p > 0.40$ ) nor a significant three-way interactions among Round, Condition, and Background variables ( $p > 0.08$ ).

Since the weak *s*-genitive condition was the only condition in which children's preference changed significantly over time, we conducted a by-subject (**Figure 8**) and by-item (**Figure 9**) analysis on this dataset. **Figure 8** shows that only two participants (subject 4 and subject 22) had decreased their proportion of *s*-genitive choice over time. Moreover, while eight participants had chosen *s*-genitive 100% of the time in the weak *s*-condition at the first round of testing, 24 participants had chosen *s*-genitive 100% of the time in the second round of testing. Only three participants (subject 20, subject 24, subject 33) in the second round of testing were at chance-level in choosing *s*-genitive, while others were above chance (either 75% or 100%).

In the item-based analysis in **Figure 9**, we can see that some items elicited more *s*-genitives than others in the first round of testing, while in the second round of testing, four out of three items (item 2: The teacher's joke/the joke of the teacher was very mean; item 3: The woman's voice/the voice of the woman was very loud; item 4: A life guard saved the man's life/the life of the man) elicited more than 90% of *s*-genitive choices. Item 1

<sup>3</sup>Since age of L2 onset (AoO) and length of residence in the L2 environment (LoE) highly correlated ( $r = -0.85$ ,  $p > 0.001$ ), we were not able to include both variables in the model due to issues of multicollinearity. Therefore, we have decided to include AoO, as suggested by the reviewers, since our sample of bilingual returnee children included children with a wide range of AoO.

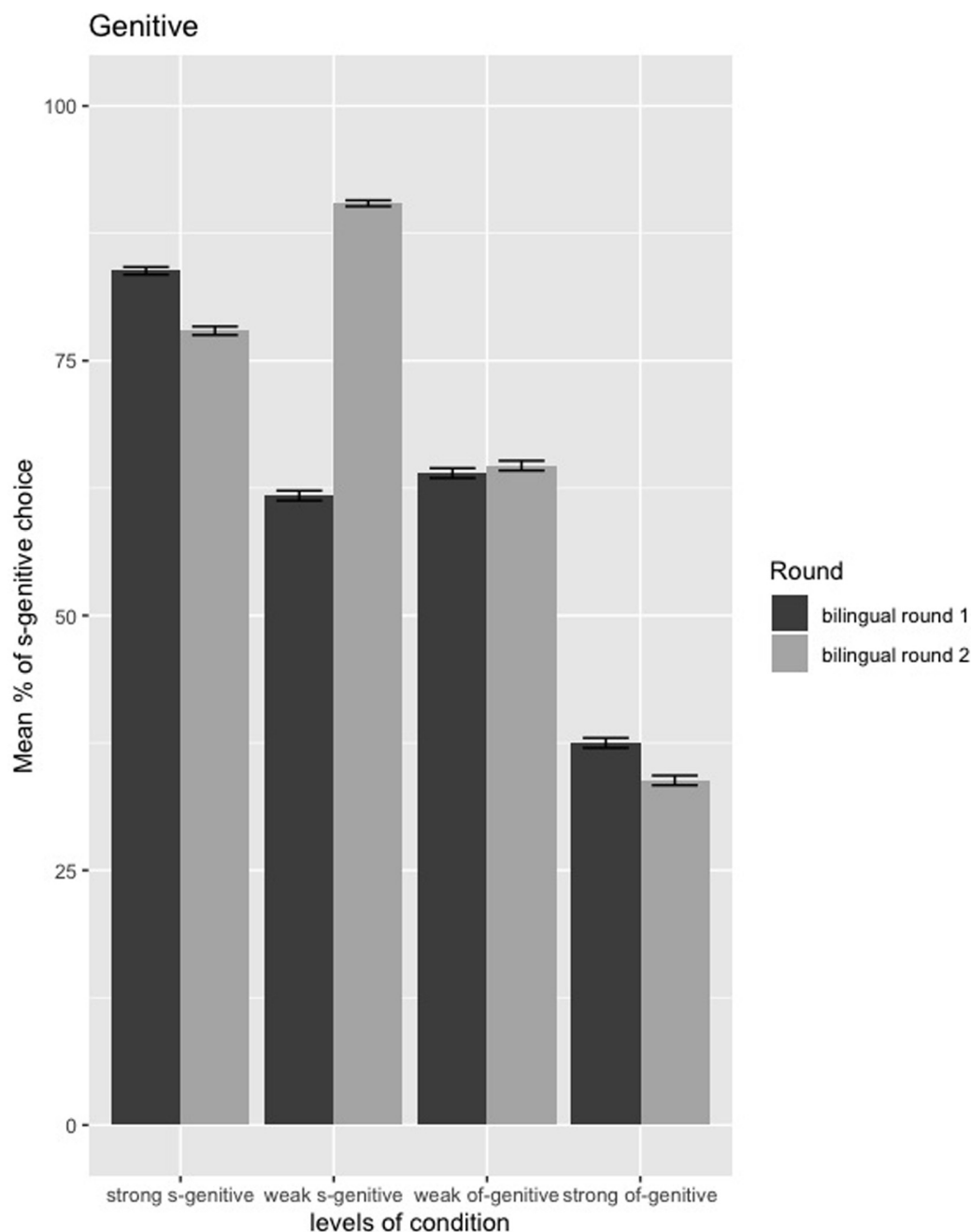


**FIGURE 7 |** Mean percentages of s-genitive choice in the genitive forced-choice task between first round and second round of testing in bilinguals; error bars = standard error.

(But nobody knew the teacher's name/the name of the teacher) was the only one that elicited less than 90% (around 82%) of s-genitive choice.

#### *Verb/argument order*

The estimated coefficients of the verb/argument order model are presented in **Supplementary Data Sheet 2: Table 4**. The model output shows that there are no significant differences within

bilinguals over time in their evaluation of verb/argument order, given the lack of significant interactions between Group and Round ( $p$ 's > 0.08). As shown in **Figure 10**, the performance on all verb/argument order conditions remained at ceiling over time (and behaved more closely to the monolinguals in the second round of testing). Despite a significant decrease in exposure over time, there appears to be no negative (or positive) changes in the returnees' evaluation of verb/argument order over time.

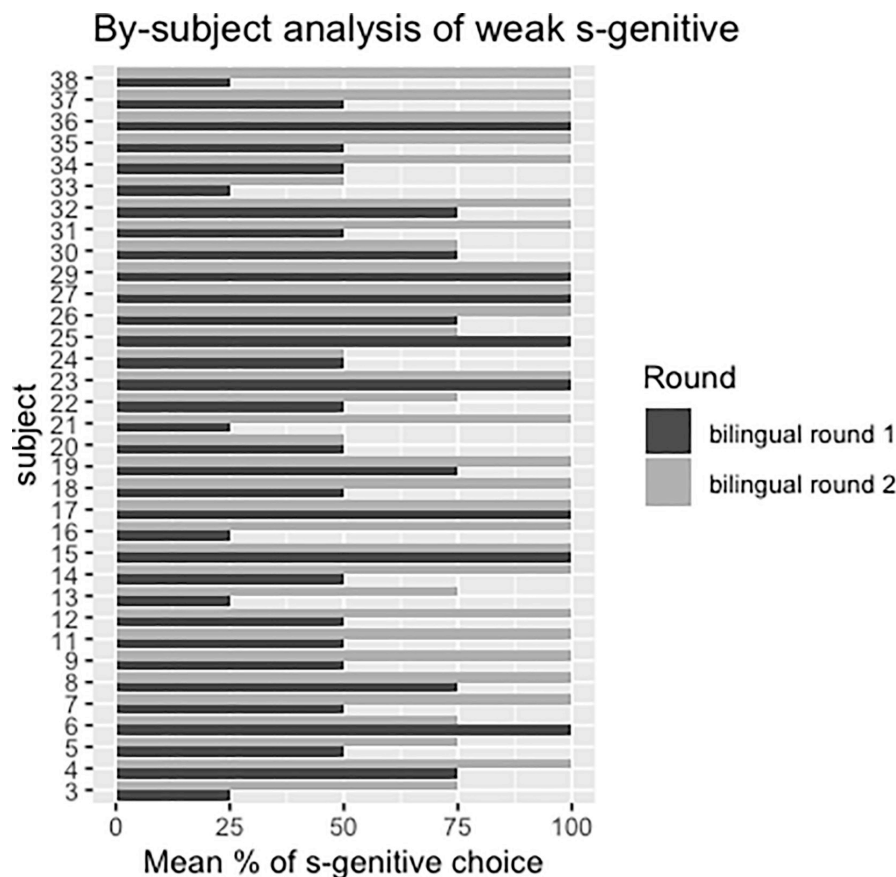


**FIGURE 8 |** Mean percentages of s-genitive choice in the weak s-genitive condition from first to second round of testing for each subject.

## Discussion

In Study 2, we examined the extent to which the bilinguals' choice of forms changes over time after the move from an L2-dominant to an L1-dominant environment. If CLI effects are driven entirely by L1 dominance, this would predict an overall increase in the acceptance of non-native-like word orders (e.g., SOV) over time, and similarly an increase of *s*-genitive choice across the genitive conditions. Our results, however, do not support this. First, there was no change in the evaluation of verb/argument order over time. Second, our data show that while the preference for *s*-genitive in the weak *s*-genitive condition did increase from first to second round of testing, it was only in this one condition that we observed an increase in *s*-genitive—there was no general increase of *s*-genitives across the board. In sum, these findings suggest that language dominance alone is not sufficient to explain our results: neither the lack of change in evaluation of verb/argument order nor the restricted change in genitive choice. What then can explain the pattern we observed? Does it instead support the other hypothesis we entertained based on a convergence of language dominance and the grammatical status of the linguistic properties tested? Recall, the prediction was that overlap and interface conditions would determine which properties would undergo increased CLI due to significant reduction in L2 input. Under this approach, verb/argument order was expected to

remain unaffected, as was borne out in the data. However, this approach also would have predicted a more generalized extension of *s*-genitive, which we did not observe. Thus, this second hypothesis was only partially confirmed, a point to which we return in greater detail below.

## GENERAL DISCUSSION

As stated earlier, the aim of this study was two-fold. First, we compared the genitive and verb/argument order choice between bilinguals and monolinguals in order to investigate whether linguistic properties play a role in the presence of CLI. Second, we tracked the genitive and verb/argument order choice of bilingual returnee children over time to test whether change in language exposure has any effect on the evaluation of these two structures.

The findings of Study 1 with respect to genitive choice revealed that bilinguals and monolinguals behaved differently on the “weak conditions” only—where prototypicality is in conflict with the other two factors (i.e., animacy and topicality). This finding is not in line with our initial predictions. Recall that we expected there to be an effect for overextending the *s*-genitive, which we predicted would manifest across all conditions. Our results

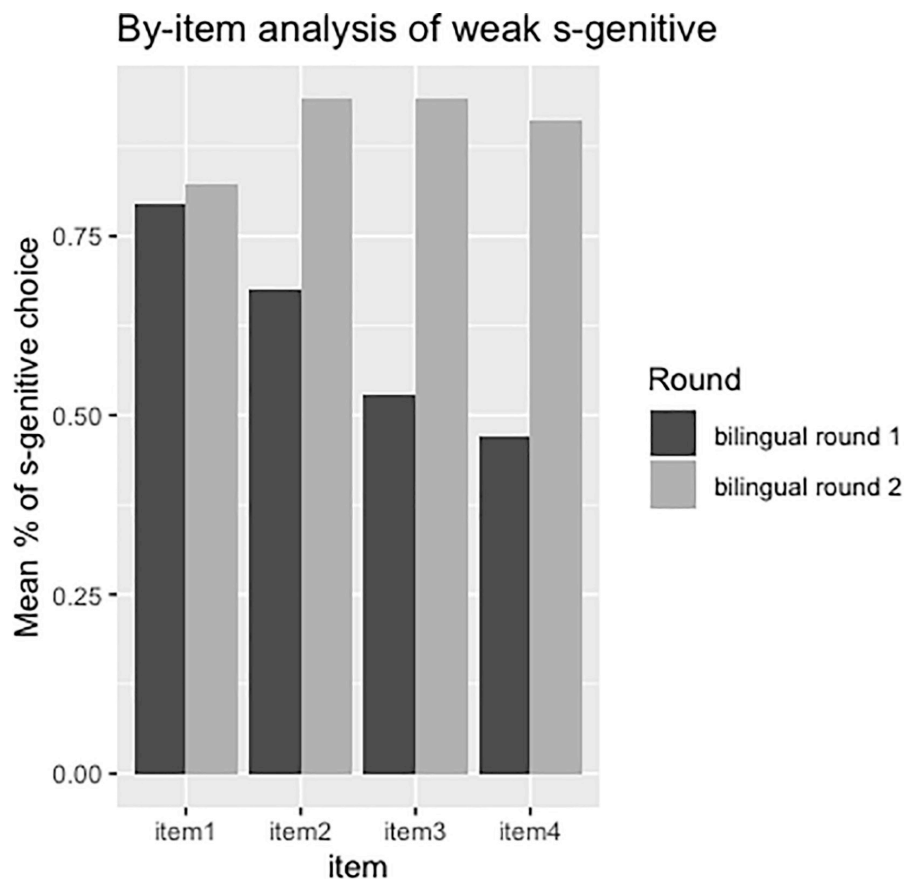


**FIGURE 9 |** Mean percentages of s-genitive choice in the weak s-genitive condition from first to second round of testing for each item.

here are instead more in line with the account proposed by Sorace (2011, 2016): namely, that the coordination of multiple factors involved in the process of choosing one structure over another is a more demanding task for bilinguals. In the current study, integrating three factors that govern the choice of genitives—animacy, topicality, and prototypicality—may be particularly taxing for bilingual children.

In the “strong” conditions, all three semantic and pragmatic factors are aligned. For example, in the strong *s*-genitive condition, *prototypicality*, *topicality*, and *animacy* are all positive valued, resulting in the strongest bias for the *s*-genitive form. When these factors are aligned, bilinguals seemingly have enough resources to process this information in a qualitatively similar way to monolinguals. As a result, monolinguals and bilinguals do not differ in choosing amongst the *s*-genitive and *of*-genitive options. In other words, the absence of conflicting information makes processing easier. However, when one factor is in conflict with the other two, the bilinguals are apparently less efficient at processing conflicting factors relative to monolinguals. The difference in the resolution of conflicting factors, then, gives rise to increased variation and affects the determinacy with which a form is selected.

As we saw in **Figures 5, 9** above, which we combine below as **Figure 11** for ease of reference, the above discussion is

supported by the distributional patterns of the bilinguals in the two weak conditions.

Two patterns can be noted: (a) in the weak *s*-genitive condition the proportion of *s*-genitive choice shows lower determinacy, gravitating close to chance than the monolingual pattern; (b) in the weak *of*-genitive condition it shows greater determinacy, moving away from chance, but in the direction of a preference opposite to that of the monolinguals. However, crucially in both cases, the pattern of the bilinguals favors the *s*-genitive, the only genitive form in English that shares structural overlap with Japanese. In fact, in both “weak” conditions bilinguals choose *s*-genitives roughly 60% of the time, as indicated in **Figure 11** below. Thus, it is not the case that the bilinguals are randomly choosing. In fact, it seems that the bilingual children are treating both “weak” conditions comparably, even though for native speakers the two “weak” conditions display essentially opposite patterns. While the bilinguals show a reduction in *s*-genitive preference in the weak *s*-genitive condition (61% for bilinguals as opposed to 80% for the monolinguals) and an increased preference for *s*-genitive in the weak *of*-genitive condition (64% for bilinguals compared to 43% for monolinguals), both are the consequence of the same underlying issue. Recall that potential influence from Japanese could only favor *s*-genitive selection. For this reason, we see that in both “weak” conditions *s*-genitive is

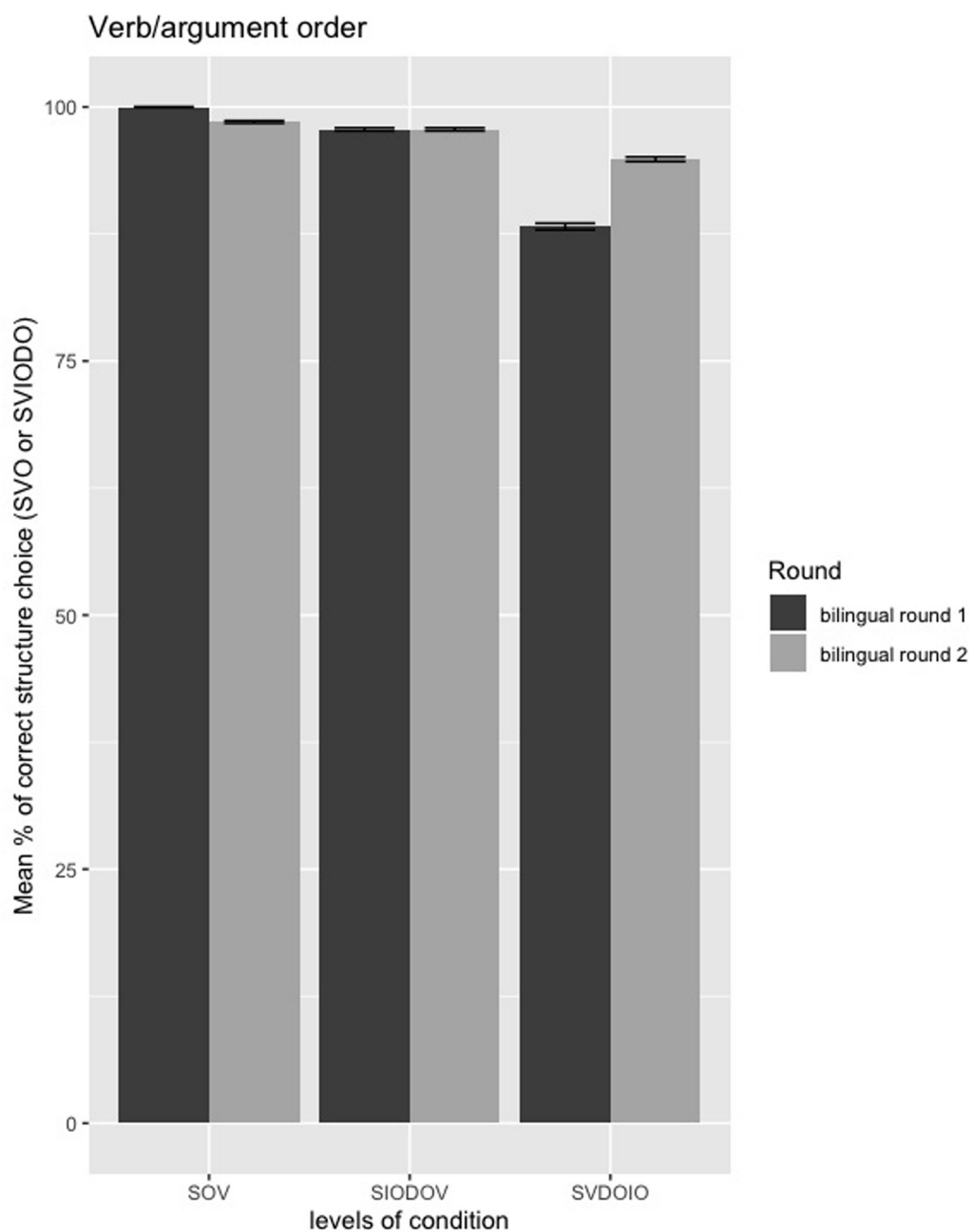


**FIGURE 10 |** Mean percentages of correct/preferred structure choice (SVO or SVIODO) in the verb/argument order forced-choice task between first and second round of testing in bilinguals; error bars = standard error.

chosen significantly above chance. However, the fact that both conditions converge toward using the *s*-genitive to the same degree suggests that conflicting cues, irrespective of how they manifest in the target grammar, are equally difficult for bilinguals. Nuance between the conditions does not obtain precisely because each has conflicting factors within the set that determines

genitive preference choice. Overall, then, increased complexity of conflicting cues creates a context in which bilinguals do not perform like monolinguals, despite clearly having a grammar with both genitive forms. In such a case, CLI effects might be attenuated in favor of a real time choice preferring *s*-genitive to the same degree, irrespective of the context.

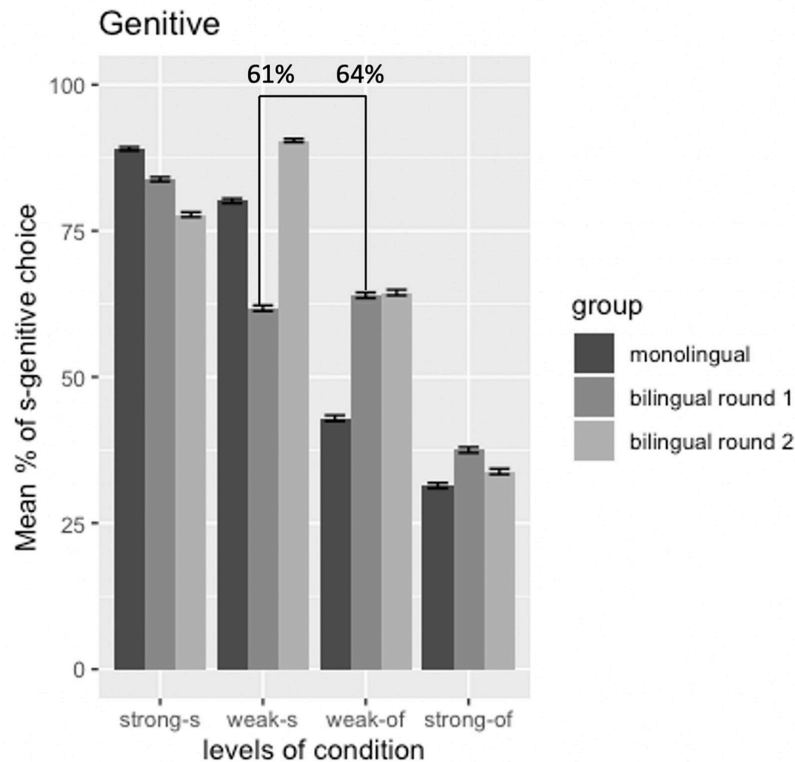


**FIGURE 11 |** Mean percentages of s-genitive choice in the genitive forced-choice task between monolinguals and bilinguals as well as bilinguals from first to second round of testing (highlighting the distributional patterns of the bilinguals in the two weak conditions); error bars = standard error.

According to the Interface Hypothesis, it is not surprising that processing conflicting factors may be more difficult for bilingual children, given that they also need to simultaneously inhibit the non-target language. In other words, allocating general cognitive resources to resolve such conflicts becomes a more demanding task, since only bilinguals have to manage the target language while suppressing the language that is not in use (Green, 1998; Meuter and Allport, 1999; Linck et al., 2009). Further, suppressing the language not in use may be particularly costly when that language is dominant, as is the case here for Japanese. This finding adds a new perspective to the hypothesis that integrating information from multiple domains increases the cognitive load (Sorace, 2011, 2016). Our results show that in addition to this, resolving conflicts between semantic and pragmatic constraints may also be cognitively demanding and thus particularly difficult to accomplish for bilingual children. We should however note that it is entirely possible that “semantic” and “pragmatic” conditions are different in terms of cognitive load, not so much because the former is “internal” and the latter is “external” *per se*, but because of the inherent higher variability of pragmatic conditions, which change in the course of contextual interaction and have to be continuously re-assessed. Our research, however, did not aim to disentangle these different conditions or measure the cognitive load they entail, and this is work that needs to be done in future research.

It may also be worthwhile in future research to investigate other cases where constraints on linguistic structures are in conflict, to examine whether there are other, similar cases of conflicting constraints where we would predict the same type of processing difficulties and behavioral corollaries in performance for bilinguals. Another obvious follow-up would be to examine genitive preference in French-English bilinguals, given that French, like Japanese, has only one genitive construction, but is the “mirror image” of Japanese in that French has only a post-nominal genitive. If processing cost, rather than CLI, explains the divergence between bilinguals and monolinguals, then French-English bilinguals should also behave differently from English monolinguals in their evaluation of genitive conditions that exhibit conflicting factors.

Turning to the verb/argument order conditions, here we found no significant differences between monolinguals and bilinguals. As discussed previously, the items from the verb/argument order conditions were formed by varying the word order of a subject, verb, and an object. This type of verb/argument order involves consideration of syntactic properties insensitive to external conditions. Processing the word order of S, V, O and IO does not require the same amount of cognitive load as genitive forms choice, which require integrating multiple semantic and pragmatic factors. Since the processing task here is relatively easy, bilinguals may have been as effective as monolinguals in parsing syntactic information.

We now turn to the question of whether the results of Study 2 are compatible with what we argued earlier on the basis of Study 1 data, or whether they call for some reconsideration. We note that the data from Study 2 are not compatible with the original hypothesis that the use of *s*-genitive increases across all four conditions. The results from Study 2 are compatible, however, with the general claims made on the basis of Study 1 data, although some qualifications are in order. Recall that in Study 2, there is a longitudinal shift in the bilinguals in one condition only, the weak *s*-genitive. That is, for all other genitive conditions, and all word order conditions, the bilinguals perform in the same way immediately on return to Japan and one year later.

We will deal with the least complicated data first, this being the word order conditions. Recall that the bilinguals in Study 1 were no different from the monolingual controls in any of the word order conditions, which we interpreted as their having acquired the target grammar in English for these core syntactic properties. The fact that there is no change in state over one year of re-immersion in their L1 Japanese in Study 2 provides rather strong evidence that core syntactic properties, once acquired, are resilient to changes in input exposure and relative language dominance in the environment.

Turning to the genitive conditions, three things are of particular significance: (a) the sharp change with the weak *s*-genitive condition, (b) the lack of change for the other weak condition (weak *of*-genitive condition) and (c) the lack of change for both “strong” conditions. The change from 61% to 90% choice of *s*-genitive in the weak *s*-genitive condition indicates clear CLI, not least because the bilinguals are now showing evidence of moving toward making this a categorical choice, going well beyond the monolingual rate of 80% in Study 1. What we see, then, is that the bilinguals’ difficulty in dealing with conflicting semantic and pragmatic cues is replaced by an even larger CLI effect. In other words, it seems that for the bilinguals there is no longer a competition between the *of* and the *s*-genitive in this condition: CLI from Japanese, in this environment of dramatically decreased exposure to the L2 and increased exposure to L1 Japanese, has led to a shift toward categorical choice of the *s*-genitive. However, if this were the whole story, why should the same effect not be evidenced in the other conditions? In other words, while this might be a reasonable explanation for (a) above, it is not enough to explain (b) and (c). Further considerations are therefore in order.

We must keep in mind that, while young in age, the bilinguals tested here were very highly proficient in their L2. After all, it was the majority language of their environment for an average of 4 years of their young lives. With this in mind, we submit that (c)—the lack of change in the “strong” conditions—is explained because there is effectively no room for Japanese CLI to be manifested in the strong *s*-genitive condition, since already at the point of return to Japan, in Study 1, the children already highly favored the *s*-genitive. For the strong *of*-genitive condition, given how proficient they are in English, we would not expect that one year would be sufficient to see a change for the extreme polar end of Rosenbach’s scale. As for (b)—the lack of change in the weak *of*-genitive condition—we should note again where this condition sits on Rosenbach’s scale (**Figure 1**). Of the three

weak *of*-genitive conditions, it is most likely to correlate with *of*-genitive choice. Therefore, we argue that again one year is not sufficient for CLI to have worked down the scale enough to see the same effect that happened for the weak *s*-genitive condition. In other words, we expect that it would take longer for Japanese CLI to overtake the bilinguals’ knowledge of English for the weak *of*-genitive condition because it is further down in Rosenbach’s scale in relation to what is most similar in form to Japanese; *s*-genitive being much closer to the Japanese *no*-genitive (linearly and in terms of its morphological status).

Recall that we presented the choice between *s*-genitive and *of*-genitive in native English in relation to Rosenbach’s scale. There are a total of eight permutations of the three semantic and pragmatic factors (see **Figure 1** above), however, we only used four conditions to make the experiments manageable for testing with children. We kept the polar ends of the scale, four conditions (two of each) in which either *s*-genitive or *of*-genitive is predicted to be most likely chosen. Effectively, we eliminated the least clear conditions, ones where the natives would be increasingly more likely to inch toward chance levels in choosing between the competing structures. Had we been able to test all the conditions, we would have been in a position to really address the hypothesis that CLI influence is a gradual process, one that follows in accord with formal descriptions such as Rosenbach’s scale. Given the language pairing and context we are working with, evidence from bilingual populations such as Japanese-English returnees could be used as a unique testing ground for determining its ecological validity. If this account is on the right track, in a situation of massively decreased exposure to English in favor of Japanese, we would expect a change in the direction of increased choice of the *s*-genitive in Japanese-English bilinguals that would be manifested first near the left end of Rosenbach’s scale and then progressively affect the other environments going from left to right; with the weak *of*-genitive condition that we tested only susceptible at a very late stage. This prediction deserves to be tested empirically in further work: this can be done by re-introducing the additional conditions on Rosenbach’s scale that had to be excluded in this study. In summary, then, we submit that the data as a whole from Study 2 suggest that re-immersion increased CLI effects, and that this happened in a gradual, yet principled way (i.e., following from a motivated linguistic hierarchy).

As pointed out by a reviewer, in order to see whether there are viable alternatives to the analysis we offered above based on conflicting semantic and pragmatic features, one could investigate the relative frequencies of the forms. That is, it could be the case that what we observed in our monolingual vs. bilingual comparison is not an effect of CLI or difficulties in conflict processing, but rather a result of delayed acquisition of genitive forms due to reduced input frequency. What we would expect from this line of reasoning is that Japanese-English bilingual children are, in fact, still in the process of acquiring the distributional patterns of English genitive forms and are following a similar (but delayed) acquisition pattern to that of typical monolinguals, precisely because the bilinguals differ from them in many aspects of language experience, including quality and quantity of input, age of

onset, and length of exposure. The significant role of input frequency in determining similar acquisition patterns has, for example, been attested in the case for Aspect as seen, for example, in Romance languages. Each of the four categories of inherent semantics of verbs or so-called *Aktionsart* (achievement, accomplishment, activity, state) can be characterized in terms of three semantic features: telic, punctual, and dynamic. Past and progressive inflections are initially restricted to certain *Aktionsart* predicates (perfective past on achievement verbs and imperfective past on stative verbs). Over time, monolingual children extend these prototypical forms to other verb types across the scale until eventually both forms can be used with all four verb types in a progressively predictable way based on viewpoint aspect (Andersen, 1991; Andersen and Shirai, 1994). And so, while initially imperfective past is used exclusively with stative verbs (-dynamic, -telic, -punctual) and perfective past with achievement verbs (+dynamic, +telic, +punctual), their prototypes, both perfective and imperfective morphology begin to be used with other verb types with non-aligning semantic features, (i.e., accomplishment and activity verbs). A similar line of reasoning could be applied to the acquisition of genitive forms in English. That is, *s*-genitive and *of*-genitive might be shown to have discernible prototypes. Their use in other contexts might be predicted to align with the progressively weaker semantic feature bundles that give rise to a preference for one or the other form, before eventually, each form can be used in all eight categories described in the Rosenbach's hierarchy.

Testing this alternative hypothesis would clearly require investigating monolingual English and Japanese-English bilingual child corpora. However, at present the existing bilingual corpora are not sufficient (see Ota, 1998 for example); further investigation would be best carried out in a follow-up study that empirically probes for what these bilingual children do in all 8 conditions (as opposed to the 4 subset conditions we tested herein). We therefore leave this work for a future, separate follow-up. We also note that, while frequency in the relevant sense could potentially have some explanatory force for the results from Study 1 (bilingual vs. monolingual comparison), it could not contribute to an explanation for the results of Study 2 where access to English input is reduced to below 5% or so. As we believe the Study 2 data are the more interesting results and, crucially, constitute the novel focus of this paper on bilingual returnees, we leave the discussion of relative frequencies of genitive forms and the acquisition patterns for future work.

## CONCLUSION

There were two objectives in our study. First, we compared relative preferences for genitive forms and verb/argument orders in English between Japanese-English bilinguals and English monolinguals, to examine whether there are any effects of cross-linguistic influence in the former. The results showed that bilinguals differed from monolinguals only in the genitive conditions, specifically in those that

required processing of semantic and/or pragmatic factors that are in conflict. These findings suggest that general processing difficulties in resolving such conflicts provide a better explanation for the observed behavior than does CLI from L1 to L2, as the Interface Hypothesis would predict. However, our results also show that not only the type of factors, but also their consistency plays a role in defining degrees of processing difficulties: therefore, it is necessary to go beyond simply contrasting “internal” and “external” interface conditions.

The second objective of our study was to investigate how severe change in continuous language input over time from the point of re-immersion in the L1 community affects returnee bilinguals' L2 grammars. We set out to establish if there were any changes in the evaluation of genitive forms and of the verb/argument orders, and if so, whether the change(s) could be explained by increased CLI effects. Our results showed that there was no change in the preference for verb/argument orders; there was a change in the preference for genitive forms over time, but that it was restricted to a single condition, namely, the weak *s*-genitive condition. In order to account for the singling out of this condition, we proposed a principled explanation for why it is most susceptible to CLI. To be clear, the rather significant change in viewpoint, as compared to our original predictions, is a direct consequence of having learned from these data themselves. Combining the results from across the two studies, we believe the data come together to nicely show the dual effect of processing complexity and influence from dominant to non-dominant language, working in tandem to explain monolingual to bilingual differences as well as longitudinal changes within bilinguals over time.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University of Edinburgh Linguistics and English Language Ethics Committee (protocol number 11–1516/5). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

MK contributed to conceptualization, methodology, validation, carried out the formal analysis, investigation, data curation, funding acquisition, wrote the original draft, and visualized the data. CH contributed to conceptualization, methodology, wrote, reviewed and edited the manuscript, and supervised the data.

AS contributed to conceptualization, wrote, reviewed and edited the manuscript, and supervised the data. JR wrote and reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

We would like to express our appreciation to the following foundations for providing financial support: Great Britain Sasakawa Foundation (Grant No. 5038), Japan Foundation Endowment Committee (Grant No. 5780416), and the British Association for Japanese Studies (Award No. JC2018-6-1). MK and JR are funded by the Tromsø Forskningsstiftelse (Tromsø Research Foundation).

## REFERENCES

- Andersen, R. W. (1991). "Developmental sequences: the emergence of aspect marking in second language acquisition," in *Crosscurrents in Second Language Acquisition and Linguistic Theories*, eds T. Huebner and C. A. Ferguson (Amsterdam: John Benjamins), 305–324. doi: 10.1075/lald.2.17and
- Andersen, R. W., and Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Stud. Second Lang. Acquis.* 16, 133–156. doi: 10.1017/s0272263100012845
- Anschutz, A. (1997). How to choose a possessive noun phrase construction in four easy steps. *Stud. Lang.* 21, 1–35. doi: 10.1075/sl.21.1.02ans
- Argyri, E., and Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Biling. Lang. Cogn.* 10, 79–99. doi: 10.1017/s1366728906002835
- Bannard, C., and Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychol. Sci.* 19, 241–248. doi: 10.1111/j.1467-9280.2008.02075.x
- Clancy, P. M. (1985). "The Acquisition of Japanese," in *The Cross-Linguistic Study of Language Acquisition*, ed. D. I. Slobin (New Jersey, NJ: Lawrence Erlbaum Associates, Inc), 373–524.
- Foroodi-Nejad, F., and Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian–English bilinguals. *Biling. Lang. Cogn.* 12, 411–427. doi: 10.1017/s1366728909990241
- Ghilzai, S. A. (2014). The role of L1 transfer in the processing of L2 English possessive constructions: evidence from German. *Eur. Academ. Res.* 2, 7454–7481.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Biling. Lang. Cogn.* 1, 67–81. doi: 10.1017/s1366728998000133
- Gries, S. T. (2002). "Evidence in linguistics: three approaches to genitives in english," in *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics?*, eds R. M. Brend, W. J. Sullivan, and A. R. Lommel (Fullerton, CA: LACUS), 17–31.
- Gries, S. T., and Wulff, S. (2013). The genitive alternation in Chinese and German ESL learners: towards a multifactorial notion of context in learner corpus research. *Int. J. Corpus Linguis.* 18, 327–356. doi: 10.1075/ijcl.18.3.04gri
- Hakuta, K. (1976). A case study of a Japanese child learning english as a second language. *Lang. Learn.* 26, 321–351. doi: 10.1111/j.1467-1770.1976.tb00280.x
- Hinrichs, L., and Szmrecsanyi, B. (2007). Recent changes in the function and frequency of standard english genitive constructions: a multivariate analysis of tagged corpora. *English Lang. Linguis.* 11, 437–474. doi: 10.1017/s1360674307002341
- Hulk, A., and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Biling. Lang. Cogn.* 3, 227–244. doi: 10.1017/s1366728900000353
- Izumi, E., and Isahara, H. (2004). "Investigation into language learners' acquisition order based on an error analysis of a learner corpus," in *Proceedings of*

## ACKNOWLEDGMENTS

We would first and foremost like to thank the parents and the children who participated in the research. Special thanks go to Japan Overseas Education Services (JOES) as well as Setagaya Children's Center for assisting us with participant recruitment and test administration.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.560874/full#supplementary-material>

- the IWLLeL 2004: an Interactive Workshop on Language e-Learning (Tokyo), 63–71.
- Jucker, A. (1993). "The genitive versus the of-construction in newspaper language," in *The Noun Phrase in English. its Structure and Variability*, ed. A. Jucker (Heidelberg: Carl Winter), 121–136.
- Kupisch, T. (2007). Determiners in bilingual German–Italian children: what they tell us about the relation between language influence and language dominance. *Biling. Lang. Cogn.* 10, 57–78. doi: 10.1017/s1366728906002823
- Leech, G., Francis, B., and Xu, X. (1994). The use of computer corpora in the textual demonstrability of gradience in linguistic categories. *Cont. Linguis. Semant.* 19, 57–76. doi: 10.1075/lis.19.07lee
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing access to the native language while immersed in a second language: evidence for the role of inhibition in second-language learning. *Psychol. Sci.* 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x
- Lozano, C. (2006). "The development of the syntax-discourse interface," in *The Acquisition of Syntax in Romance Languages*, eds V. Torrens and L. Escobar (Amsterdam: John Benjamins), 371–399.
- Luk, Z. P., and Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? evidence from the acquisition of plural-s, articles, and possessive's. *Lang. Learn.* 59, 721–754. doi: 10.1111/j.1467-9922.2009.00524.x
- Margaza, P., and Bel, A. (2006). Null subjects at the syntax-pragmatics interface: evidence from Spanish interlanguage of Greek speakers. 88–97.
- Meuter, R. F., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Memory Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602
- Müller, N., and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Biling. Lang. Cogn.* 4, 1–21. doi: 10.1017/s1366728901000116
- Nicoladis, E. (2002). What's the difference between 'toilet paper' and 'paper toilet'? French-English bilingual children's crosslinguistic transfer in compound nouns. *J. Child Lang.* 29, 843–863. doi: 10.1017/s0305000902005366
- Nicoladis, E. (2012). Cross-linguistic influence in French–English bilingual children's possessive constructions. *Biling. Lang. Cogn.* 15, 320–328. doi: 10.1017/s1366728911000101
- Nuibe, Y. (1986). A report on the development of the grammatical morphemes in Japanese junior high school students learning English as a foreign language. *Kyoiku Kagaku [Educational Science]* 28, 371–381.
- Ota, M. (1998). "Minimality constraints and the prosodic structure of child Japanese," in *Japanese/Korean Linguistics*, Vol. 8, ed. D. Silva (Stanford, CA: CSLI Publication), 331–344.
- Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *Int. J. Biling.* 5, 19–38. doi: 10.1177/13670069010050010201
- Paradis, J., and Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: what is the role of the input? *J. Child Lang.* 30, 371–393. doi: 10.1017/s03050009030005609

- Pullum, G. K. (1977). Word order universals and grammatical relations. *Syntax Semant.* 8, 249–277. doi: 10.1163/9789004368866\_011
- Rosenbach, A. (2001). “The English s-genitive: animacy, topicality, and possessive relationship in a diachronic perspective,” in *Historical Linguistics 1999*, ed. L. Brinton (Amsterdam: John Benjamins), 277–292. doi: 10.1075/cilt.215.19ros
- Rosenbach, A. (2002). *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies*. New York, NY: Mouton de Gruyter.
- Rosenbach, A. (2003). “Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English,” in *Determinants of Grammatical Variation in English*, eds G. Rohdenburg and B. Mondorf (New York, NY: Mouton de Gruyter), 379–412.
- Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language* 81, 613–644. doi: 10.1353/lan.2005.0149
- Rosenbach, A. (2014). English genitive variation—the state of the art. *Engl. Lang. Linguist.* 18, 215–262. doi: 10.1017/S1360674314000021
- Rosenbach, A., and Vezzosi, L. (2000). “Genitive constructions in Early Modern English: New evidence from a corpus analysis,” in *Stability, Variation, and Change of Word-order Patterns Over Time*, eds R. Sornicola, E. Poppe, and A. Shisha-Halevy (Amsterdam: John Benjamins), 285–307. doi: 10.1075/cilt.213.20ros
- Rothman, J., Alonso, J. G., and Puig-Mayenco, E. (2019). *Third Language Acquisition and Linguistic Transfer*. Cambridge: Cambridge University Press.
- Serratrice, L., Sorace, A., Filiaci, F., and Baldo, M. (2009). Bilingual children’s sensitivity to specificity and genericity: evidence from metalinguistic awareness. *Biling. Lang. Cogn.* 12, 239–257. doi: 10.1017/s1366728909004027
- Shirahata, T. (1988). The learning order of English grammatical morphemes by Japanese high school students. *JACET Bull.* 19, 83–102.
- Skarabela, B., and Serratrice, L. (2009). “The doctor’s mother’ or ‘the mother of the doctor’?: syntactic priming of possessive noun phrases in English preschoolers,” in *Online Supplement of the Proceedings of the Boston University Conference on Language Development* (Boston: Boston University).
- Snap, N., Matthews, J., Hirakawa, M., Hirakawa, Y., and Hosoi, H. (2014). Aspect in L2 English: a longitudinal study of four Japanese child returnees. *EUROSLA Yearbook* 14, 79–110. doi: 10.1075/eurosla.14.04sna
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches Biling.* 1, 1–33. doi: 10.1075/lab.1.1.01sor
- Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguistic Approaches Biling.* 6, 669–684. doi: 10.1075/lab.15055.sor
- Sorace, A., and Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Lang. Res.* 22, 339–368. doi: 10.1191/0267658306sr271oa
- Sorace, A., and Serratrice, L. (2009). Internal and external interfaces in bilingual language development: beyond structural overlap. *Int. J. Biling.* 13, 195–210. doi: 10.1177/1367006909339810
- Stefanowitsch, A. (2003). “Constructional semantics as a limit to grammatical alternation: the two genitives of English,” in *Determinants of Grammatical Variation in English*, eds G. Rohdenburg and B. Mondorf (New York, NY: Mouton de Gruyter), 413–444.
- Szmrecsanyi, B., and Hinrichs, L. (2008). “Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres,” in *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, eds T. Nevalainen, I. Taavitsainen, P. Pahta, and M. Korhonen (Amsterdam: John Benjamins), 291–309. doi: 10.1075/silv.2.22szm
- Teramura, H. (1991). *Syntax and Semantics in Japanese*. Tokyo: Kurosio Publishers.
- Tomiyama, M. (1999). “The first stage of second language attrition: A case study of a Japanese returnee,” in *Second Language Attrition in Japanese Contexts*, ed. L. Hansen (Oxford: Oxford University Press), 59–79.
- Tomiyama, M. (2000). Child second language attrition: a longitudinal case study. *Appl. Linguistics* 21, 304–332. doi: 10.1093/applin/21.3.304
- Treffers-Daller, J., and Silva-Corvalán, C. (2016). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press. doi: 10.1093/applin/21.3.304
- Unsworth, S. (2003). Testing Hulk and Müller (2000) on crosslinguistic influence: root infinitives in a bilingual German/English child. *Biling. Lang. Cogn.* 6, 143–158. doi: 10.1017/s1366728903001093
- Unsworth, S. (2016). “Amount of exposure as a proxy for dominance in bilingual language acquisition,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173. doi: 10.1017/cbo9781107375345.008
- White, L. (2011). The interface hypothesis: how far does it extend? *Linguistic Approaches Biling.* 1, 108–110. doi: 10.1075/lab.1.1.19whi
- Wolk, C., Bresnan, J., Rosenbach, A., and Szmrecsanyi, B. (2013). Dative and genitive variability in Late Modern English: exploring cross-constructional variation and change. *Diachronica* 30, 382–419. doi: 10.1075/dia.30.3.04wol
- Yip, V., and Matthews, S. (2000). Syntactic transfer in a Cantonese–English bilingual child. *Biling. Lang. Cogn.* 3, 193–208. doi: 10.1017/s136672890000033x
- Yoshitomi, A. (2007). “Testing the primacy of aspect and reverse order hypothesis in Japanese returnees: Towards constructing a corpus of second language attrition data,” in *Corpus-based Perspectives in Linguistics*, eds Y. Kawaguchi, T. Takagaki, N. Tomimori, and Y. Tsuruga (Amsterdam: John Benjamins Publishing), 371–389. doi: 10.1075/ubli.6.25yos

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kubota, Heycock, Sorace and Rothman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Individual Chunking Ability Predicts Efficient or Shallow L2 Processing: Eye-Tracking Evidence From Multiword Units in Relative Clauses

Manuel F. Pulido\*

Department of Spanish, Italian and Portuguese, Center for Language Science, Penn State University, University Park, PA, United States

## OPEN ACCESS

### Edited by:

Tamar Degani,  
University of Haifa, Israel

### Reviewed by:

Kathy Conklin,  
University of Nottingham,  
United Kingdom  
Henrik Gyllstad,  
Lund University, Sweden

### \*Correspondence:

Manuel F. Pulido  
mpulido@psu.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 17 September 2020

**Accepted:** 21 December 2020

**Published:** 15 January 2021

### Citation:

Pulido MF (2021) Individual  
Chunking Ability Predicts Efficient or  
Shallow L2 Processing: Eye-Tracking  
Evidence From Multiword Units in  
Relative Clauses.  
Front. Psychol. 11:607621.  
doi: 10.3389/fpsyg.2020.607621

Behavioral studies on language processing rely on the eye-mind assumption, which states that the time spent looking at text is an index of the time spent processing it. In most cases, relatively shorter reading times are interpreted as evidence of greater processing efficiency. However, previous evidence from L2 research indicates that non-native participants who present fast reading times are not always more efficient readers, but rather shallow parsers. Because earlier studies did not identify a reliable predictor of variability in L2 processing, such uncertainty around the interpretation of reading times introduces a potential confound that undermines the credibility and the conclusions of online measures of processing. The present study proposes that a recently developed modulator of online processing efficiency, namely, chunking ability, may account for the observed variability in L2 online reading performance. L1 English – L2 Spanish learners' eye movements were analyzed during natural reading. Chunking ability was predictive of overall reading speed. Target relative clauses contained L2 Verb-Noun multiword units, which were manipulated with regards to their L1-L2 congruency. The results indicated that processing of the L1-L2 incongruent units was modulated by an interaction of L2 chunking ability and level of knowledge of multiword units. Critically, the data revealed an inverse U-shaped pattern, with faster reading times in both learners with the highest and the lowest chunking ability scores, suggesting fast integration in the former, and lack of integration in the latter. Additionally, the presence of significant differences between conditions was correlated with individual chunking ability. The findings point at chunking ability as a significant modulator of general L2 processing efficiency, and of cross-language differences in particular, and add clarity to the interpretation of variability in the online reading performance of non-native speakers.

**Keywords:** chunking ability, individual differences, eye-tracking, L2 proficiency, multiword units, second language, processing, generalization

## INTRODUCTION

For several decades, psycholinguistic studies on first (L1) and second language (L2) processing have employed measures of reading times as an indicator of ease of processing. This connection rests on the “eye-mind assumption” (Just and Carpenter, 1980), i.e., the notion that the amount of time spent reading a word is representative of the time spent processing it. Although the

eye-mind association is now known to be more complex than previously held (Miller, 2015), the basic rationale remains valid and is the foundation of a vast body of literature focused on processing (for methodological reviews of L2 processing studies see, e.g., Frenck-Mestre, 2005; Dussias, 2010; Jiang, 2015; Keating and Jegerski, 2015; Marsden et al., 2018). Longer reading times in psycholinguistic studies are typically associated with processing costs induced by relatively more difficult conditions; while faster reading times are associated with relative ease of processing in less demanding conditions.

However, in spite of these robust and well-documented effects, previous evidence has indicated that faster reading does not always index processing efficiency (Broughton et al., 2010; Kaan et al., 2015; Miller, 2015). For example, readers often take fast-reading strategies that favor “good-enough” interpretations (Karimi and Ferreira, 2016; Metzner et al., 2017), even if this comes at the expense of misinterpreting the input (e.g., sentences with non-canonical syntactic structures) (Ferreira et al., 2001; Ferreira, 2003; Ferreira and Patson, 2007). While more engaged individuals may also read faster (Broughton et al., 2010), fast reading times may be simply due to lack of engagement or to a good-enough approach to process information. This point has become perhaps even more evident with recent methodological advances. In a recent experiment, Metzner et al. (2017) examined fixation-related brain potentials time-locked to eye movements during natural reading. The co-registration data in Metzner et al. (2017) provide perhaps the first direct evidence of shallow processing at the neurophysiological level that is coupled with faster-than-expected reading behavior.

As in studies with native speakers, there is good evidence that L2 readers may sometimes take similar speed-favoring strategies even when this compromises comprehension and the ability to process the input effectively, resulting in shallow processing (Felser et al., 2003; Roberts and Felser, 2011; Kaan et al., 2015). The evidence of shallow reading in L2 speakers led Clahsen and Felser (2006) to propose the shallow structure hypothesis (SSH). The SSH suggests that L2 users may strategically focus on lexical, pragmatic and other surface-level cues to achieve efficient processing that may be on par with L1 performance (Clahsen and Felser, 2006). Others (e.g., Hopp, 2010, 2013) have specifically proposed that using the L2 poses greater demands on the cognitive system and limits the resources available when using the less-dominant language. For example, the lack of automatization in lexical retrieval during early stages of processing may cascade into difficulties in forming syntactic representations in real time (Hopp, 2015). This latter perspective may also account for shallowness of processing in some of the participants of L2 processing studies, who fail to show the effects of experimental manipulations in comparisons between conditions. While it is not the goal of this paper to adjudicate between the different accounts, it does make a strong prediction about the availability of cognitive resources as a factor critically modulating individual L2 processing performance.

When reading in a non-native language, processing costs may be exacerbated as a result of lower L2 proficiency (Kotz and Elston-Güttler, 2004; Jeon and Yamashita, 2014), as well as lack of syntactic-semantic congruency between the L1 and L2

(Tokowicz and MacWhinney, 2005; Foucart and Frenck-Mestre, 2011; Wolter and Gyllstad, 2011), and slower lexical access (Hopp, 2013, 2015), among other issues. While such difficulties are usually associated with longer reading times, this is not always the case. Two important findings from previous research are that (a) strikingly, some L2 users may read even faster than control native speakers, exhibiting implausibly short reading times that indicate shallow reading (e.g., Experiments 2 and 4 in Felser et al., 2003; Kaan et al., 2015); and that (b) when matched in proficiency, faster non-native readers may present smaller differences between conditions and reduced grammaticality effects, suggesting shallower parsing (Kaan et al., 2015).

If faster reading times can index processing that is deep, engaged and efficient, but also processing that is shallow, disengaged and inefficient, it is clear that this creates an important confound which critically affects the interpretation of reading times in psycholinguistic studies<sup>1</sup>. It is still unknown what specific factors predict when some readers will show faster or slower reading times in the L2. Language proficiency has long been known to be a strong predictor of reading comprehension in an L2 (Alderson, 1984, 1993; Bowey, 1986; Koda, 2005; Kieffer and Lesaux, 2008, 2012a,b; Shiotsu, 2010; for a recent meta-analysis see Jeon and Yamashita, 2014). While lower linguistic skills in non-native speakers are typically associated with slower reading times, especially when compared with L1 speakers (Coughlin and Tremblay, 2012), online measures have yielded mixed evidence on whether reading speed correlates with proficiency (Roberts and Felser, 2011; Kaan et al., 2015). In other words, there is a lack of understanding of the modulators driving the strategic deployment of resources during reading, which may be able to account for and predict L2 reader performance. Clarifying the underlying modulators of L2 processing should be a priority, if potential confounds are to be avoided in the conclusions drawn from reading experiments. However, in spite of these findings, most studies have continued to straightforwardly consider faster reading times as a hallmark of efficiency in processing.

A specific cause of concern is the possibility that, in at least some cases, such low-efficiency readers may make up a non-negligible proportion of the sample. If correct, this implies that shallow reading is, at the very least, a likely contributor of noise in the data and a cause of a potential confound in the conclusions reached. That is, linguistic and cognitive measures are believed to modulate reading times but, based on previous evidence, I propose that this relationship may be non-linear: Individuals with higher L2 and/or cognitive skills should tend to present smaller differences between experimental conditions as well as faster reading times. However, a similar pattern is expected from individuals with low cognitive resources and low proficiency, if they only engage with the input at a shallow level. This may produce an inverse U-shaped curve, with relatively faster reading times at each end of the processing-efficiency continuum, even if for entirely different reasons.

<sup>1</sup>Importantly, this fact is still fully coherent with the eye-mind assumption, in that less time spent looking at a word still indicates less time processing that word, even if the underlying cause is, e.g., lack of engagement.

To address this hypothesis, the present research examines whether individual differences in linguistic and cognitive abilities that support online processing may account for the variability in reading speed among L2 readers. To do so, it considers the role of L2 lexical knowledge in conjunction with chunk sensitivity, i.e., a recently developed cognitive measure found to be a significant predictor of processing efficiency, and a modulator of online reading (McCauley and Christiansen, 2015; McCauley et al., 2017; López-Beltrán et al., 2020). While a number of well-known measures of cognitive skill have been investigated, their predictive power in what concerns online processing appears to be limited. For example, greater engagement of executive control has been found to be associated with efficiency in recovery from initial misinterpretations, both in the L1 (Novick et al., 2013; Hsu and Novick, 2016) and in the L2 (Navarro-Torres et al., 2019). However, the role of executive control in such studies is most theoretically relevant in cases in which conflicting representations require controlled selection, specifically, rather than in processing across the board. Other previous studies have often focused on the role of working memory (WM) in reading. There is ample evidence that WM capacity predicts outcomes in offline reading comprehension both in the first (Baddeley, 1979; Daneman and Carpenter, 1980; Siegel, 1994) and in the second language (Harrington and Sawyer, 1992; Payne et al., 2009; Sagarra and Herschensohn, 2010; Erçetin and Alptekin, 2013; Hopp, 2014). However, WM appears to play a small role in measures of online L2 processing (Juffs, 2004; Havik et al., 2009; Roberts, 2012; Hopp, 2014; López-Beltrán et al., 2020; although see Navarro-Torres et al., 2019).

Recent evidence has suggested that better chunking ability is associated with more efficient online processing in native (McCauley and Christiansen, 2015; McCauley et al., 2017) and non-native speakers (López-Beltrán et al., 2020). Briefly, previous work has proposed that in order to deal with the immediacy of language, speakers must be sensitive to the structural probabilities in the input, if they are to successfully process the linguistic signal in real time (“the now-or-never bottleneck”; Christiansen and Chater, 2016). Based on well-known cognitive constraints that limit the amount of input information that can be maintained in memory (Miller and Taylor, 1948; Elliott, 1962; Pashler, 1988; Remez et al., 2010), Christiansen and Chater (2016) proposed that chunking may play a critical role in facilitating real-time processing, by allowing humans to recode the incoming signal into chunks at multiple levels of abstraction (from phonemes, to words, phrases and sentences). To illustrate, while recalling a sequence such as *h c r l t i a p a c e a p* poses a considerable challenge, the task becomes easier when the same string is rearranged into a sequence of recognizable words, as in *c a t a p p l e c h a i r*. Given previous claims that non-native speakers may not always process the input efficiently in real time, the role of chunk sensitivity may be a particularly relevant predictor of online L2 comprehension. The present study will assess the role of chunking ability as a potential predictor of online processing and reading speed. To do this, measures of chunk sensitivity will be collected from L1 English—L2 Spanish learners using the tasks available in each of those two languages. If, as suggested by some accounts (Hopp, 2006, 2010; Kaan et al., 2015), differences

in the resources available to L1 and L2 speakers are responsible for reading patterns, chunking ability may indicate, as an index of L2 processing efficiency, whether relevant linguistic cues are in fact processed online. An important question, however, is whether L2 processing is best predicted by chunk sensitivity measured in the L1 or the L2. So far, this question has only been explored in one self-paced reading study (López-Beltrán et al., 2020). Using eye-tracking, the present study will investigate the potential effect of individual chunking ability as online L2 sentence processing unfolds.

## THE PRESENT STUDY

To investigate individual variability in L2 reading, this study will capitalize on processing costs induced by multiword units that are incongruent across the native and non-native language, as a tool to investigate the processing of information that is particular to the L2. It therefore builds on a considerable number of studies that have identified, within the last decade, L1-L2 incongruent multiword units as a locus of processing costs (e.g., Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013; Carrol et al., 2016; Wolter and Yamashita, 2018; Pulido, 2020).

Multiword units that are congruent (i.e., have word-by-word equivalents) are known to experience a processing advantage (even when encountered for the first time, e.g., Carrol and Conklin, 2017). On the other hand, L1-L2 incongruent multiword units, which differ at least in part from their L1 counterparts, are notoriously difficult to acquire (e.g., Nesselhauf, 2003; Laufer and Girsai, 2008; Pulido and Dussias, 2020) (e.g., in Spanish *pedir una hamburguesa* is equivalent to “order a hamburger,” but it literally translates as “request a hamburger”). A number of studies have consistently found cross-linguistic costs in processing of L1-L2 incongruent multiword units such that, even when these are well known, they produce costs in L2 processing (Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013; Wolter and Yamashita, 2018). Therefore, L1-L2 incongruent collocations provide an ideal testbed to investigate variability in L2 reading based on L2 experience and individual differences in processing efficiency. In the present study, participants will be presented with multiword units composed of a verb and a noun. In addition, while the previous studies investigating cross-language effects on collocational processing employed reaction time tasks (such as lexical decision or phrase acceptability judgments), the experiment reported here examines online processing during natural reading.

An eye-tracking reading task is employed to examine learners’ processing of L2 multiword units, half of which are L1-L2 incongruent, e.g., *pedir una hamburguesa* (“order a hamburger”). The sentences included a relative clause in which the verb-noun unit was reversed, so that the incongruent element (i.e., the verb) would be focalized after the noun (e.g., *Él dice que las hamburguesas que pedirán son las mejores de la ciudad*, “He says that the hamburguers they will order are the best in town”). While previous data indicates that whether a multiword unit is presented in the canonical or a non-canonical order affects its recognition speed (e.g., in binomials;

Siyanova-Chanturia et al., 2011b), in the present design this manipulation affects all target items equally. Importantly, eye-tracking data has indicated that the properties of multiword units are retained even when reversed (e.g., an idiom advantage is shown by “the bucket was kicked”, based on the canonical “kick the bucket”; Kyriacou et al., 2020). In this line, the L1-L2 congruency effect of collocations is expected to remain unaffected. Thus, data from eye movements collected during natural reading will allow to investigate the effect of individual-based differences in processing efficiency and speed during natural reading, in conditions of high ecological validity.

Based on previous findings, in the present study I hypothesize that online L2 reading is modulated by individual linguistic and cognitive skills. Specifically, in the approach taken here, rather than dividing the sample into overall slow/fast readers as in some previous studies (e.g., Hoover and Dwivedi, 1998; Roberts and Felser, 2011; Kaan et al., 2015), L2 multiword knowledge and individual chunking ability are examined as potential factors driving differences in reading times.

## Research Questions

RQ1. In what concerns processing efficiency, how do chunking ability measures (in the L1 and/or the L2) modulate reading times?

RQ2. In connection with L2 proficiency, how does command of L2-specific multiword knowledge modulate reading measures?

RQ3. Do chunking ability and L2 proficiency (indexed by L2 multiword unit knowledge) interact to modulate reading times and, if so, in what manner?

RQ4. In what concerns cross-linguistic differences, how do these measures differ or converge during processing of L2 multiword units that are congruent with the L1 equivalents, or L1-L2 incongruent?

RQ5. Based on eye-tracking measures, at what stage of processing will individual differences in processing become apparent? Specifically, how do L2 multiword knowledge and chunking ability affect eye movements in early measures (i.e., gaze duration) and late measures (i.e., total reading times)?

## Predictions

If faster reaction times in previous studies are associated with more efficient processing in some readers, but also with shallow processing in individuals with lower chunking ability, one would expect an inverse U-shaped pattern in the effect of chunking ability on reading speed (RQ1). If this prediction is correct, the research questions will clarify the contribution of each factor to modulate reading times of L2 multiword units, when these are congruent and incongruent with their L1 counterparts. Regarding the two versions of the Chunk Sensitivity task (in the L1 and the L2), previous results from self-paced reading data suggest that L2 processing might be predicted by chunking ability measured in the L1.

Regarding the role of L2 proficiency, the evidence from previous studies is mixed (RQ2). On the one hand, higher proficiency should be correlated with faster reading speed. At lower proficiency levels, slower reading times would be typically expected. However, knowledge of L2-specific features may interact with chunking ability in complex ways (RQ3). In particular, individuals at the low end of both L2 multiword units and chunking ability may be particularly prone to engage with the input at a shallow level.

The prediction for RQ4 is that, in reading V-N phrases, items that L1-L2 incongruent should result in greater costs (e.g., Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013). More specifically, cross-language effects should emerge at elements that differ across the L1 and L2 (e.g., the verb in phrases such as “las hamburguesas que pedirán,” as in the example above). But no cross-language effects should emerge at the nouns, which are congruent across the L1 and the L2.

Finally, at least one previous study has found that both early and late measures were sensitive to the degree of conventionalization of V-N multiword units (Vilkaitė, 2016). Given important differences between the present study and the stimuli of previous studies, no specific predictions are made in what concerns early versus late measures of processing, or their possible interaction with the other variables investigated. Thus, the measures of gaze duration and total duration will be examined with no strong *a priori* expectations.

## MATERIALS AND METHODS

### Participants

A group of 45 participants was recruited at The Pennsylvania State University. Participants were native speakers of English who were enrolled in third and fourth semester university Spanish courses (roughly equivalent to level B1 of the Common European Framework of Reference for Languages, Council of Europe, 2011). This sample size is comparable to that of previous studies that have consistently detected cross-language effects during processing of L2 multiword units (Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013), and is in line with recent studies that examined individual-based differences in chunking ability, both in the L1 (McCauley et al., 2017) and the L2 (López-Beltrán et al., 2020)<sup>2</sup>.

Participants completed a linguistic background questionnaire as well as proficiency measures to confirm they met the required proficiency level. One subject who was identified as an early bilingual was excluded. Five additional participants were excluded due to accuracy lower than 70% during the reading task ( $N = 1$ ), due to experimental error during data acquisition ( $N = 1$ ), or because they failed to complete all the sessions ( $N = 3$ ). The results and data analysis reported below are based on the remaining thirty-nine participants (77% female). The study was approved by the Institutional Review Board; all participants

<sup>2</sup>An “observed power” estimation performed with the *simr* package in R (v. 1.0, Green and MacLeod, 2016) indicated that a sample of  $N = 25$  was sufficient to reach 80% power,  $\alpha = 0.05$ .

gave informed consent and were paid 10 US dollars per hour of participation.

## Materials and Procedure

In a first session, participants completed individual differences measures in this order: English Chunk Sensitivity, Spanish Vocabulary test, Spanish Chunk Sensitivity task and Language History Questionnaire; L2 tasks were blocked this way to avoid switching repeatedly between the L1 and L2, while separating the two versions of the Chunk Sensitivity task. Participants then completed three sessions as part of a multiword unit learning study that will be reported elsewhere (Pulido, 2020). Two weeks after the first session, participants returned to the lab to complete the Multiword Units test and the Reading task. This way, multiword unit knowledge was tested immediately before the main Reading task. The sequencing of tasks is illustrated in **Figure 1**.

For ease of presentation, this section first describes the materials created for the Reading task, followed by the materials of individual differences tasks employed to measure chunking ability and language proficiency.

## Reading Task Materials and Procedure

### List of V-N multiword units

First, a list of thirty-eight target V-N phrases was created, half of which were congruent (i.e., had literal equivalents) across English and Spanish, while the other half were cross-linguistically incongruent, i.e., specific to the L2. While phrases congruent with the native language are interpretable based on cross-language similarity (Siyanova-Chanturia et al., 2011a; Pulido and Dussias, 2020), incongruent items require knowledge specific to the L2. Given this, incongruent items were selected based on their similarity to a list of collocations that had been recently learned by the participants. For example, based on the collocation *pedir pizza* (“request pizza,” equivalent to “order pizza”) previously learned by the participants, the related V-N phrase *pedir una hamburguesa* (“order a hamburger”) was created. This made all target phrases employed here interpretable based on prior knowledge.

Furthermore, to ensure the interpretability of incongruent multiword units, the semantic relatedness between the nouns in the previously learned and the current incongruent target phrases was measured through PMI (pointwise mutual information) (Budi et al., 2007; Bullinaria and Levy, 2007; Recchia and Jones, 2009). PMI has been shown to be a valid index of semantic relatedness, and to even be more highly correlated to human

judgments than more computationally intensive measures such as Latent Semantic Analysis, when based on large corpora (> 400 million words, Recchia and Jones, 2009)<sup>3</sup>. Data from a large web-based corpus of over two billion words (Corpus del Español, Davies, 2016) was used to calculate the scores of similarity and confirm the semantic similarity between the nouns in previously learned and target phrases. To compare related target pairs to an unrelated baseline, the similarity scores of target pairs were compared against a list of unrelated pairs, consisting of the same words re-matched (e.g., *tren* – *pizza*, “train” – “pizza”). This confirmed that target items were semantically similar to known items (mean similarity score for fillers:  $-2.06$ ; SD:  $1.23$ ; mean for related items:  $2.63$ , SD:  $2.33$ ;  $p < 0.0001$ ). The full set of phrases (including the related previously learned incongruent items) are available in the **Supplementary Material**.

Importantly, in all incongruent phrases, the verb was the element that differed across the L1 and the L2, while the noun was literally congruent. Lists of congruent and incongruent phrases were matched on properties including the length and log frequency of verb and the noun; per-million observations of verb-noun multiword units; and verb-noun association strength (t-scores). Given that the focus was on processing of the verbs, and in order to minimize comprehension problems in reading novel phrases, nouns in all V-N phrases were cognates.

### Reading materials

Based on each multiword unit from the list described above, a preamble and a carrier sentence were created, which were followed by a comprehension question. The preamble was included to provide context, and the ensuing carrier sentence contained the target multiword unit. The length of preambles and target carrier sentences was not significantly different for trials containing congruent and incongruent multiword units (all  $p > 0.50$ ). Half of the comprehension questions items (50% true, 50% false) were related to the content of the preamble and the other half to the carrier sentence. A full trial is illustrated in **Table 1**. The full list is available in the **Supplementary Material**.

As indicated, all units were composed of a verb and a noun. However, carrier sentences included a relative clause in which

<sup>3</sup>An additional advantage of PMI is its computational simplicity, which allows to develop semantic relatedness measures for more comprehensive lists of words, including words without very high frequencies.

**TABLE 1** | Sample materials for one experimental trial.

### Preamble

Juan planea invitar a algunos amigos a su restaurante favorito  
 (“Juan plans to invite some friends to his favorite restaurant”)

### Target sentence

Él dice que las hamburguesas que *pedirán* son las mejores de la ciudad  
 (“He says that the hamburgers that they *will order* are the best in town”)

### Comprehension question

¿Él recomienda la fruta?

(“Does he recommend the fruit?”) (Response: False)

*In these examples, target multiword units appear underscored, with verbs shown in italics (no text enhancement was used during the experiment).*

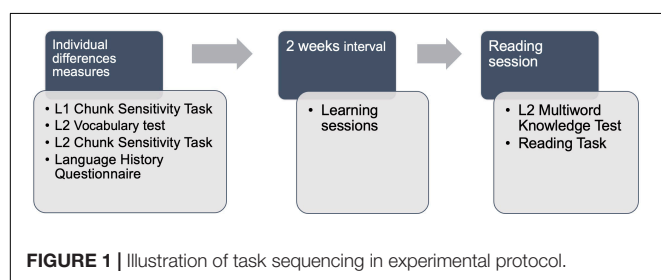


**FIGURE 1** | Illustration of task sequencing in experimental protocol.

the V-N phrase was reversed, so that the incongruent element (i.e., the verb) would follow the noun. To illustrate, based on the V-N phrase *pedir hamburguesas* “order hamburgers,” a carrier sentence was created: *Él dice que las hamburguesas que pedirán son las mejores de la ciudad* (“He says that the hamburgers that they will order are the best in town”). Given that Spanish and English have virtually identical syntax in subject relative clauses, the syntactic frame itself is cross-linguistically congruent (i.e., Spanish N-“*que*”-V, English “N-[*that*]-V”)<sup>4</sup>. Further, although as in English, subject relative clauses are “non-canonical” relative to the V-N order in Spanish, they are very highly productive. For example, while “*pedir*” + “*hamburguesa*” in the V-N order occurs 131 times in the Corpus del español (4 word span), the inverted N-V order also occurs 17 times, in a ratio of about 1 to 8.

This manipulation served several goals. First and foremost, given the focus on verbs (which are L1-L2 incongruent in half of the items), the relative sentence structure allows to present a verb that is maximally predictable based on the preceding context. That is, unlike in a V-N order of presentation, seeing the noun (direct object) first means that the verb is fully interpretable as soon as it is encountered, relative to cases in which the verb’s object is presented later in the sentence. Secondly, because participants had previously learned verb-noun collocations in the canonical word order as part of a larger project, this context allows to examine the role of chunking ability independently from prior practice with a given syntactic frame. Third, it extends previous research which examined the effect cross-language congruency when reading V-N phrases to sentences containing relative clauses.

### Reading task procedure

Participants’ eye movements were recorded using an Eyelink 1000 plus eye-tracker (SR Research, Inc). Eye movements were recorded from the right eye. Participants were seated at 90 cm from the monitor and comfortably rested their chins on a chin rest. At the start of the experiment, a nine-point calibration procedure was performed, followed by a calibration accuracy test. Calibration was checked at the beginning of each trial, and was repeated if any point had an error greater than 1°, or if the average error for all points was greater than 0.5°. Each sentence was presented left-aligned in the center of the screen. Text was displayed in Consolas size 15 font. To begin each trial, participants looked at a fixation point coinciding with the first character position of the sentence. Participants were instructed to read each sentence at their own pace, and to press a button to advance to the next screen. The preamble and target sentences appeared in separate screens in one single line; therefore, the target multiword phrase was never displayed at the beginning or end of the line. At the end of the trial, participants responded to the question by pressing a “yes” or a “no” button on a hand-held device. Four practice trials preceded the experimental items. The task lasted approximately 25 min.

<sup>4</sup>In English the “that” complementizer is optional; it is always expressed in Spanish, making the relative structure explicit. Besides other structural differences (i.e., English requires overt subjects, but these are optional in Spanish) sentence have equivalent syntax across languages, as in the examples in Table 1.

## Individual Differences Measures

Participants completed a measure of Chunk Sensitivity in both of their languages (L1-English and L2-Spanish) as well as additional measures to assess short-term memory and language proficiency in the L1 and L2. The measures are described below and their outcomes are reported in the section “Results.”

### Chunk sensitivity tasks

Participants’ individual chunking ability was measured using two versions of the Chunk Sensitivity task, each in one of participants’ languages (L1 English and L2 Spanish). In this task, participants are instructed to recall strings of 12 individual words, each made up by 4 trigrams. The task includes 20 strings, evenly divided into target and control trials. Each target trial consists of four frequent trigrams extracted from native speaker corpus data, which lend themselves to be “chunked” as a unit (e.g., have to eat good to know don’t like them is really nice). Matched controls are made up of unrelated words with no statistical association (e.g., years got don’t to game have she mean to them far is), and thus provide a baseline for short-term memory span. A chunk sensitivity index is calculated as the difference in recall accuracy between targets and controls.

Responses were recorded and coded offline. Each correctly recalled word in a string is awarded 1 point, for a maximum of 3 points per trigram (12 points for the whole string). One point is deducted from the whole trial for imperfect ordering.

The English version employed here contained the materials in the original task reported in McCauley et al. (2017), which was created based on data from the American National Corpus (Reppen et al., 2005) and the Fisher corpus (Cieri et al., 2004). The Spanish version was developed by López-Beltrán et al. (2020), based on data from the *Corpus del español* (Davies, 2016); the complete description of the tasks is reported in López-Beltrán et al. (2020).

### Phonological short-term memory

While the chunking ability tasks are designed to factor out the role of phonological short-term memory (PSTM) from recall of multiword chunks, previous research has shown that PSTM is on its own a significant predictor of learners’ ability to retrieve L2 collocations (Pulido and Dussias, 2020). Therefore, participants were administered a Non-word Repetition task (Baddeley et al., 1988) as an index of PSTM. The lists employed here were based on the materials reported in Martin and Ellis (2012). Four lists of three, four, five or six non-words were presented in ascending order. Participants’ responses were recorded and were scored offline following a scheme adapted from Gathercole et al. (2001). One full point was awarded for each correctly recalled non-word (up to a maximum of 72). For partially correct non-words, 0.25 was deducted for each error in the position of a phoneme; and 0.5 was deducted for each missing phoneme, or for phonemes that were not part of the trial.

### Language experience and proficiency measures

To assess the linguistic profile and background in the L1 and L2, participants completed the LEAP-Q (Marian et al., 2007), which contained items related to weekly usage and exposure to the L2.

### Vocabulary test

Prior word knowledge was measured through a multiple-choice vocabulary test. While knowledge of L2 multiword units was important to the goals of the study and measured separately, the rationale for this test was to assess participants' knowledge of the basic meaning of individual verbs and nouns, outside the scope of L2 multiword units. Items consisted of the basic meaning of 84 items, including the nineteen verbs used in the L1-L2 incongruent multiword units; for these units, the basic word meaning assessed in the test (e.g., “pedir” = “request”) differs from its specialized meaning in an L2-specific collocation (as described, “pedir hamburguesas” literally translates as “request hamburgers” but is equivalent to “order hamburgers”). Therefore, the test allowed to gauge individual word knowledge, as opposed to knowledge of multiword units; this was considered important because a learner might know the basic meaning of the word, even if the specialized meaning is less familiar<sup>5</sup>. The test materials are available in the **Supplementary Material**.

### L2 multiword knowledge test

A L2 multiword unit test measured participants' knowledge of previously learned incongruent V-N collocations as part of a larger project (reported in Pulido, 2020), and was administered right before the reading experiment [see section “Reading Task Materials and Procedure” on how the current materials were related to the learned items]. The L2 multiword knowledge test, which followed previous learning of L1-L2 incongruent collocations as part of a larger study, provided a baseline individual measure of L2-specific multiword knowledge. Participants were presented with L1 verb-noun phrases and were asked to provide the previously learned L2 translation. To clarify, it tested multiword unit knowledge by using items which were not included in the present study. Knowledge of multiword units is the critical aspect of L2 proficiency under examination, and this measure is referred to as “L2 multiword knowledge.”

## Data Cleaning and Analysis

The reading measures reported are gaze duration and total reading times. The analysis reported here follows recent studies on reading of Verb-Noun collocations that examined processing using a combination of early (gaze duration) and late measures (total duration) (e.g., Vilkaite, 2016; Vilkaite and Schmitt, 2019). Gaze duration is defined as the sum of all eye-fixations on the critical region of interest before leaving it the first time that it is read; total times are the sum of all fixations on the critical region, including regressions (re-reading). As described, each Verb-Noun phrase (e.g., *pedir hamburguesas* “order hamburgers”) was presented in a relative sentence, e.g., *las hamburguesas que pedirán* “the hamburgers that they will order.” Gaze and total durations were extracted for reading times of the verb (in the previous example, *pedirán* “they will order”), where a stronger effect was expected; as well as for the preceding noun

phrase (*las hamburguesas* “the hamburgers”), which served as a baseline. While no effects were expected in the noun region for gaze duration (i.e., for fixations made prior to reading the verb), potential effects might be present for the measure of total duration, which includes regressions to the noun (e.g., after reading the verb).

Only trials with correct comprehension responses were included in the analysis. One item in the incongruent condition was excluded due to experimental error<sup>6</sup>. All participants included in the analysis met the threshold of 70% accuracy in comprehension (mean: 87.82% correct responses); one participant with low comprehension accuracy that was close to chance (57% correct) was removed. Inspection of the data revealed that, in transitioning from the preamble sentence to the target sentence, fixations were made on various parts of the sentence in some trials. To avoid repetition effects due to foveal and parafoveal processing of target regions in such cases, trials in which the first fixation was not on the first word of the sentence were removed (7.15%).

Total durations shorter than 100 ms or longer than 3000 ms were excluded (5.07%). The data were z-scored, and outliers were then removed for each condition based on their individual median absolute deviation (MAD) for each participant and condition. The MAD method is a more robust measure for outlier removal than standard deviations, given that the latter are susceptible to be distorted by observations that strongly deviate from the mean (Leys et al., 2013). Trials above or below 3 absolute deviations from the median were excluded, resulting in 3.27% data loss.

Separate analyses were conducted for each of the two dependent measures extracted (gaze duration and total duration times) for the verb and for the determiner-noun. That is, four separate models (2 measures  $\times$  2 interest areas) were developed following the same procedure. Data skewness in the dependent variable was corrected by log-normalizing the data<sup>7</sup>.

Mixed-effects modeling was conducted using the lme4 package (version 1.1-23; Bates et al., 2015) in R (version 4.0.2). Model fitting always started with a core model which included English and Spanish Chunk Sensitivity scores, L1-L2 Congruency and Collocation test scores (“L2 multiword knowledge”), to address the main research questions. Then the potential interaction and additional covariates were added to the model, which included: L1-L2 Congruency (congruent, incongruent), individual scores for L2 vocabulary and phonological short-term memory. Additionally, stimulus properties were considered, including log frequency and collocational strength (i.e., t-scores) of multiword units; as well as orthographic length and log

<sup>6</sup>One item displayed an erroneous carrier sentence; excluding this item did not affect the results of the analyses.

<sup>7</sup>Measures of data skewness revealed that total duration times were right-skewed (skewness for total duration of noun region: 1.50 and the verb region: 1.60); after being log-normalized, skewness was corrected (noun region: -0.10; verb region: 0.11). The gaze duration data was not right-skewed, i.e., unlike total reading times, they did not contain small numbers of observations with very high reading time values (noun region: 0.14; verb region: 0.37). Because applying a log transformation was not necessary to correct skewness and resulted in artificially non-normally distributed gaze duration data (skewness for noun region: -2.82; verb region: -2.08), gaze duration was not log-transformed.

<sup>5</sup>While the individual word test employed here is highly relatable to the materials employed, it may also be limited in terms of its generalizability and characterization of participants' overall level of proficiency. To overcome this limitation, future work should incorporate other normed tests (e.g., the Spanish version of the Lextale test; Izura et al., 2014).

frequency of the individual words in the region of interest. For models of the verb, the same variables were considered, but orthographic length and frequency of the preceding noun were also included. All continuous variables were centered. Variance Inflation Factors were calculated using the *vif* function of the *car* package (version 3.0-9; Fox et al., 2007), and indicated no substantial collinearity among the variables considered (all VIF < 2).

The random effects structure included by-subject and by-item random intercepts, as well as random by-subject slopes for trial number and L1-L2 Congruency, and by-item slopes for English and Spanish Chunk Sensitivity and previous Collocation test scores. Following convergence issues with the maximally specified structure (Barr et al., 2013), the random effects structure was simplified. All final models included by-subject and by-item random intercepts, as well as by-subject random slopes for trial number.

Starting from this full model, a backward step-by-step model selection process was adopted. Variables were removed one by one, starting with those with the lowest *t*-values. Predictors that did not significantly improve the model fit (likelihood ratio test  $p > 0.05$ ) were removed. The code of each of the models is provided in the **Supplementary Material**. The following section reports on the results of the selected models. The results presented include 95% confidence intervals (CI) and parameter-specific *p*-values estimated using the normal approximation.

## RESULTS

### Baseline Measures

The results of all the individual-based measures are reported below in **Table 2**. The scores from the baseline multiword units provided a broad proficiency range (range 42.9–96.4%), adequate for the goal of assessing the role of multiword-based proficiency along with chunking ability in processing. L1 and L2 Chunk Sensitivity were not significantly correlated ( $r = 0.15$ ,  $p = 0.37$ ). Similarly, neither measure was significantly correlated with the Multiword Units test scores (all *p*-values > 0.18). Cronbach's alpha was calculated as a reliability index and found to be acceptable for all the tests (i.e., ranging from 0.70 to 0.90; Streiner, 2003; Tavakol and Dennick, 2011), including the vocabulary test ( $\alpha = 0.75$ ), the L2 multiword knowledge test ( $\alpha = 0.80$ ), the PSTM test ( $\alpha = 0.86$ ) and the Chunk Sensitivity tasks in the English ( $\alpha = 0.86$ ) and Spanish ( $\alpha = 0.81$ ) versions.

### Reading Times of the Noun Region

#### Gaze Duration (Noun Region)

As expected, the analysis of gaze duration for the noun revealed no significant effects of Chunk Sensitivity in the L1-English measure ( $\beta: -12.76$ , SE: 13.88, CI:  $-39.11, 13.59$ ,  $p = 0.36$ ), nor in the L2-Spanish version of the task ( $\beta: 15.69$ , SE: 13.81, CI:  $-10.45, 41.92$ ,  $p = 0.26$ ). There was also no significant effect of individual knowledge of L2 collocations ( $\beta: 3.77$ , SE: 13.55, CI:  $-22.22, 29.87$ ,  $p = 0.78$ ). Finally, as expected, cross-language Congruency did not significantly affect gaze duration of the noun

**TABLE 2 |** Summary of cognitive and proficiency measures.

|                                 | Valid N | M     | SD    | Range   |
|---------------------------------|---------|-------|-------|---------|
| Age (in years)                  | 39      | 18.76 | 0.85  | 18–21   |
| Weekly exposure to L2 (%)       | 39      | 6.26  | 4.74  | 0–18    |
| Baseline Vocabulary Test (/10)  | 39      | 8.64  | 0.61  | 7.1–9.8 |
| L1 Chunk Sensitivity            | 39      | 37.72 | 12.43 | 10–58   |
| L2 Chunk Sensitivity            | 39      | 13.72 | 7.72  | –2–39   |
| PSTM: Non-word repetition (/10) | 37      | 6.12  | 0.99  | 3.7–8.0 |
| L2 multiword knowledge (/10)    | 39      | 7.88  | 1.34  | 4.3–9.6 |

*All scores reported are based on a scale from 0 to 10, unless otherwise indicated. The Chunk Sensitivity measures are the difference scores between target and control trials (with up to 120 points in each condition).*

baseline ( $\beta: 24.21$ , SE: 19.27, CI:  $-13.75, 62.25$ ,  $p = 0.21$ ). The results are shown in **Table 3**.

#### Total Duration (Noun Region)

The results of total duration for the noun region indicated no significant effect of Chunk Sensitivity in the L1 ( $\beta: 0.03$ , SE: 0.04, CI:  $-0.04, 0.11$ ,  $p = 0.38$ ). There was also no significant effect of L2 multiword knowledge ( $\beta: 0.02$ , SE: 0.04, CI:  $-0.06, 0.09$ ,  $p = 0.66$ ), nor of L1-L2 Congruency ( $\beta: 0.12$ , SE: 0.08, CI:  $-0.04, 0.27$ ,  $p = 0.12$ ). However, there were significant effects of L2 Chunk Sensitivity, with higher Spanish chunking ability reducing total duration times ( $\beta: -0.08$ , SE: 0.04, CI:  $-0.15, -0.005$ ,  $p < 0.05$ ).

### Reading Times of the Verb Region

#### Gaze Duration (Verb Region)

The analysis of gaze duration for the verb revealed a significant effect of L1 Chunk Sensitivity, such that greater chunking ability was associated with faster reading times ( $\beta: -28.05$ , SE: 12.77, CI:  $-52.56, -3.60$ ,  $p < 0.05$ ). However, L2 Chunk Sensitivity did not significantly influence the dependent measure ( $\beta: 18.06$ , SE: 12.21, CI:  $-5.16, 41.18$ ,  $p = 0.14$ ). There was a highly significant effect of L1-L2 Congruency, with slower reading times for incongruent trials relative to congruent trials ( $\beta: 58.38$ , SE: 17.97, CI:  $23.02, 93.83$ ,  $p < 0.01$ ). Finally, gaze durations were influenced by the knowledge of L2 collocations and by vocabulary scores. Interestingly, these effects went in different directions, such that greater multiword knowledge was associated with longer gaze durations ( $\beta: 58.38$ , SE: 17.97, CI:  $13.02, 61.17$ ,  $p < 0.05$ ), while higher knowledge of individual words reduced reading times ( $\beta: -27.61$ , SE: 13.04, CI:  $-52.17, -3.06$ ,  $p < 0.05$ ); these differences are further commented on in the discussion. The results are summarized in **Table 4**. The effect of verb cross-language congruency is illustrated in **Figure 2**.

#### Total Duration (Verb Region)

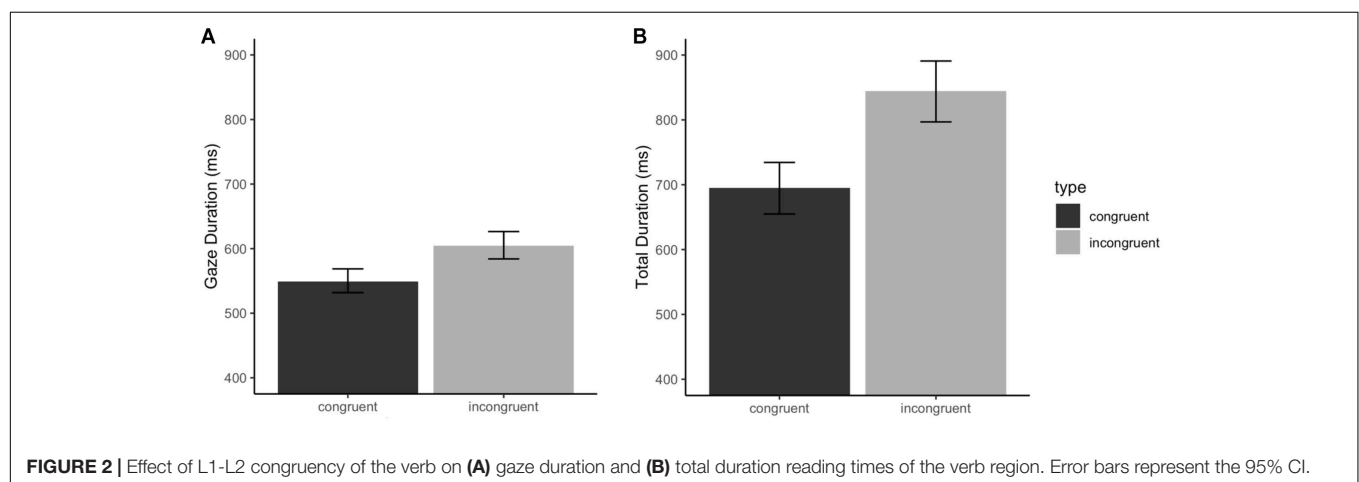
Similarly to the results of gaze duration, the results of total durations for the verb revealed a highly significant effect of L1-L2 Congruency ( $\beta: 0.22$ , SE: 0.07, CI:  $0.09, 0.36$ ,  $p < 0.001$ ). The analysis indicated no simple effect of Chunk Sensitivity in the L1 ( $\beta: -0.04$ , SE: 0.05, CI:  $-0.13, 0.06$ ,  $p = 0.44$ ) or in the L2 ( $\beta: -0.07$ , SE: 0.05, CI:  $-0.16, 0.03$ ,  $p = 0.19$ ). Importantly, a crucial significant three-way interaction emerged between L2

**TABLE 3 |** Summary of selected models for the noun region.

|                                                                                  | Gaze duration |           |              |          | Total duration                                                                  |           |              |          |
|----------------------------------------------------------------------------------|---------------|-----------|--------------|----------|---------------------------------------------------------------------------------|-----------|--------------|----------|
| Fixed Effects                                                                    | $\beta$       | <i>SE</i> | <i>t</i>     | <i>p</i> | $\beta$                                                                         | <i>SE</i> | <i>t</i>     | <i>p</i> |
| (Intercept)                                                                      | 585.39        | 17.12     | 34.19        | ***      | 6.42                                                                            | 0.06      | 100.18       | ***      |
| Eng. (L1) Chunk Sensitivity                                                      | −12.76        | 13.88     | −0.92        | 0.36     | 0.03                                                                            | 0.04      | 0.89         | 0.38     |
| Span. (L2) Chunk Sensitivity                                                     | 15.69         | 13.81     | 1.14         | 0.26     | −0.08                                                                           | 0.04      | −2.07        | *        |
| L2 multiword knowledge                                                           | 3.77          | 13.55     | 0.28         | 0.78     | 0.02                                                                            | 0.04      | 0.45         | 0.66     |
| Congruency (incongruent)                                                         | 24.21         | 19.27     | 1.26         | 0.21     | 0.12                                                                            | 0.08      | 1.55         | 0.12     |
| Random effects                                                                   | Variance      | <i>SD</i> | Correlations |          | Variance                                                                        | <i>SD</i> | Correlations |          |
| Intercept   Participant                                                          | 3542.7        | 59.52     |              |          | 0.05                                                                            | 0.22      |              |          |
| Trial Number   Participant                                                       | 187.1         | 13.68     | 0.11         |          | 0.01                                                                            | 0.12      | 0.34         |          |
| Intercept   Item                                                                 | 596.5         | 24.42     |              |          | 0.05                                                                            | 0.21      |              |          |
| Marginal <i>R</i> <sup>2</sup> : 0.007, Conditional <i>R</i> <sup>2</sup> : 0.06 |               |           |              |          | Marginal <i>R</i> <sup>2</sup> : 0.03, Conditional <i>R</i> <sup>2</sup> : 0.32 |           |              |          |
| <i>*p</i> < 0.05, <i>**p</i> < 0.01, <i>***p</i> < 0.001.                        |               |           |              |          |                                                                                 |           |              |          |

**TABLE 4 |** Summary of selected models for the verb region.

|                                                      | Gaze duration |           |              |                                                 | Total duration |           |              |          |
|------------------------------------------------------|---------------|-----------|--------------|-------------------------------------------------|----------------|-----------|--------------|----------|
| Fixed Effects                                        | $\beta$       | <i>SE</i> | <i>t</i>     | <i>p</i>                                        | $\beta$        | <i>SE</i> | <i>t</i>     | <i>p</i> |
| (Intercept)                                          | 545.92        | 15.56     | 35.07        | ***                                             | 6.33           | 0.06      | 99.3         | ***      |
| Eng. (L1) Chunk Sensitivity                          | −28.05        | 12.77     | −2.20        | *                                               | −0.04          | 0.05      | −0.77        | 0.44     |
| Span. (L2) Chunk Sensitivity                         | 18.06         | 12.21     | 1.48         | 0.14                                            | −0.07          | 0.05      | −1.31        | 0.19     |
| L2 multiword knowledge                               | 36.99         | 12.34     | 3.00         | **                                              | 0.07           | 0.05      | 1.46         | 0.14     |
| Congruency (incongruent)                             | 58.38         | 17.97     | 3.25         | **                                              | 0.22           | 0.07      | 3.42         | ***      |
| L2 Vocabulary                                        | −27.61        | 13.04     | −2.12        | *                                               |                |           |              |          |
| L2 Chunk. × L2 Colloc.                               |               |           |              |                                                 | 0.00           | 0.06      | 0.06         | 0.95     |
| L2 Chunk. × Congruency                               |               |           |              |                                                 | 0.07           | 0.03      | 2.06         | *        |
| L2 Chunk. × L2 Colloc. × Cong.                       |               |           |              |                                                 | −0.09          | 0.04      | −2.37        | *        |
| Random Effects                                       | Variance      | <i>SD</i> | Correlations |                                                 | Variance       | <i>SD</i> | Correlations |          |
| Intercept   Participant                              | 3219.65       | 56.74     |              |                                                 | 0.07           | 0.27      |              |          |
| Trial Number   Participant                           | 36.85         | 34.34     | 0.04         |                                                 | 0.01           | 0.09      | 0.00         |          |
| Intercept   Participant                              | 1179.28       |           |              |                                                 | 0.03           | 0.18      |              |          |
| Marginal $R^2$ : 0.04, Conditional $R^2$ : 0.11      |               |           |              | Marginal $R^2$ : 0.05, Conditional $R^2$ : 0.30 |                |           |              |          |
| <i>*p &lt; 0.05, **p &lt; 0.01, ***p &lt; 0.001.</i> |               |           |              |                                                 |                |           |              |          |



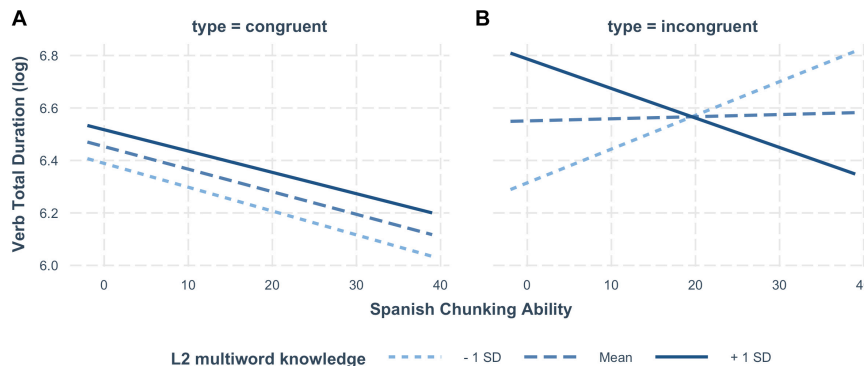


**FIGURE 3 |** Illustration of the effect of three-way interaction between on (A) congruent and (B) incongruent multiword units. L2 multiword knowledge serves as the critical L2 proficiency measure.

Chunk Sensitivity, L2 multiword knowledge and Congruency ( $\beta$ :  $-0.09$ , SE:  $0.04$ , CI:  $-0.17$ ,  $-0.02$ ,  $p < 0.05$ ). The two-way interaction between L2 Chunk Sensitivity and Congruency was also significant ( $\beta$ :  $0.07$ , SE:  $0.03$ , CI:  $0.06$ ,  $0.14$ ,  $p = 0.07$ ). The *interactions* package in R (Long, 2019) was used to visualize and aid in interpreting these effects, which are illustrated in **Figure 3**; the interaction is examined in depth in the discussion. For congruent collocations, the data presented no interaction between the main effects of L2 chunking ability and multiword knowledge. Although higher chunking ability had a facilitatory effect, it is relevant to note that reading times were shorter for readers with less proficient L2-multiword knowledge. **Table 4** summarizes the model output.

For incongruent collocations, critically, L2 chunking ability and multiword knowledge interacted, giving rise to an approximately inversed U-shape trend, such that the faster reading times were found in individuals with either low multiword knowledge and low chunk sensitivity, or with high multiword knowledge coupled with high chunk sensitivity. Interestingly, as can be observed in **Figure 3**, participants with lower multiword knowledge presented faster reading times, which became increasingly slower as chunking ability increased; for participants with average multiword knowledge, chunk sensitivity did not appear to strongly influence reading behavior; finally, individuals with higher multiword knowledge showed faster reading durations as chunking ability increased.

#### Follow-up analysis

Previous studies have performed a median split based on overall reading times, with the goal of further characterizing the individual variability within samples of L2 readers (e.g., Hoover and Dwivedi, 1998; Roberts and Felser, 2011; Kaan et al., 2015). However, as noted, in the present data on incongruent collocations faster reading times were found at both ends of the proficiency (i.e., L2 multiword knowledge) and chunk sensitivity measures. The results are indicative of the predicted inverse U-shaped trend at the sample level, where readers with higher and lower chunking ability present faster reading times, albeit for entirely different reasons. This is what is shown by the pattern illustrated in **Figure 4**; the division based on chunking ability

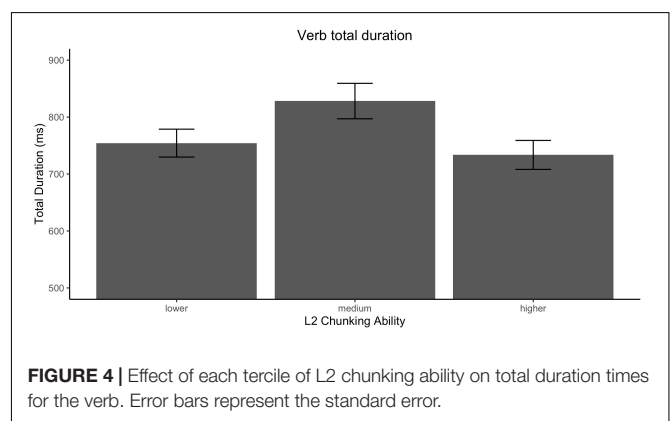


**FIGURE 4 |** Effect of each tercile of L2 chunking ability on total duration times for the verb. Error bars represent the standard error.

confirms the prediction that high- and low-chunking ability readers have similar reading speed. As a result, a follow-up analysis that divided participants based on overall reading times might lump together high- and low-efficiency readers.

To further investigate the effect of L2 multiword knowledge and its interaction with Spanish Chunk Sensitivity, *post hoc* *t*-tests (FDR-corrected) were performed to gauge the effect of L1-L2 congruency within each subgroup of participants, based on chunking ability (high, medium, and low) and L2 multiword knowledge (i.e., high or low L2). The results in **Table 5** provide further evidence that differences between conditions were driven by chunking ability, rather than knowledge of the L2 units. Crucially, high chunking ability readers presented a significant effect of congruency, regardless of L2 knowledge. On the contrary, readers with low chunking ability did not show congruency effects, also regardless of L2 knowledge.

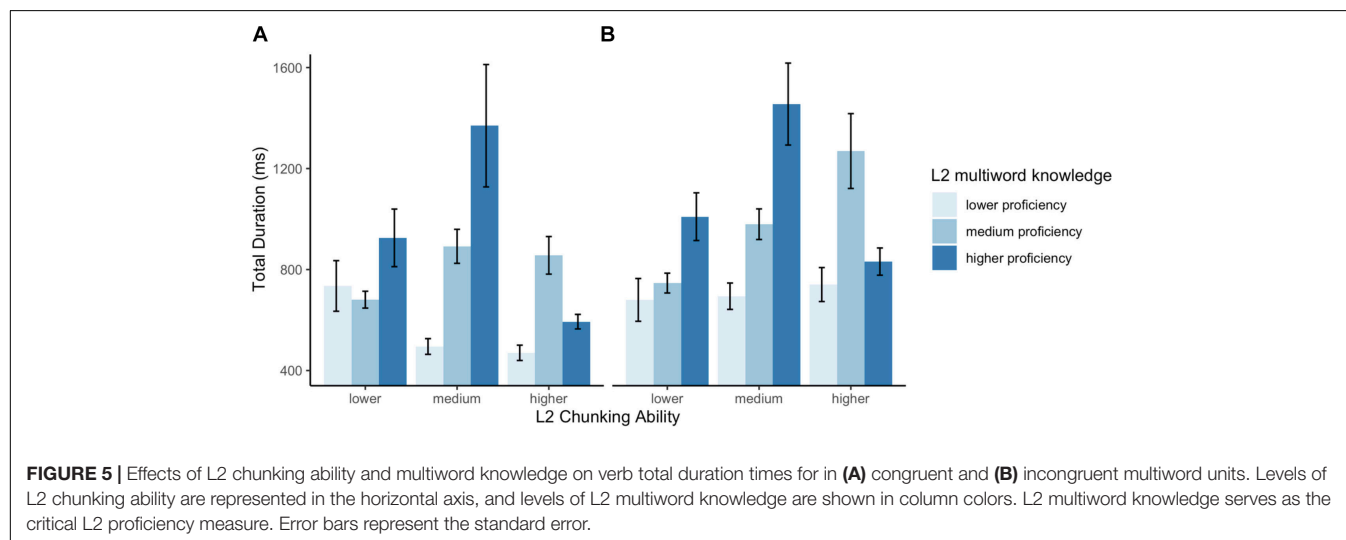
By way of visual illustration, **Figure 5** further presents each variable split into the lower, medium and higher terciles (chunking ability is shown in the horizontal axis, with columns for each L2 tercile)<sup>8</sup>. Three main points should be noted. First, an inverse-U shape pattern can be observed for readers with high knowledge of L2-specific multiword units (darker columns),

<sup>8</sup>A tercile split follows the same procedure as a median split, but the sample is divided into three quantiles instead of two (lower, medium and upper).

**TABLE 5 |** Results of the dedicated congruency-based pairwise comparisons.

| L2 multiword knowledge | Chunking ability (tercile) | Congruent mean (ms) (SD) | Incongruent mean (ms) (SD) | df     | t     | N | 95% CI       | p    |
|------------------------|----------------------------|--------------------------|----------------------------|--------|-------|---|--------------|------|
| Low                    | Low                        | 661 (474)                | 699 (504)                  | 122.27 | −0.61 | 5 | −0.28, 0.15  | 0.55 |
| Low                    | Medium                     | 624 (472)                | 822 (550)                  | 236.93 | −3.54 | 8 | −0.46, −0.13 | **   |
| Low                    | High                       | 614 (349)                | 919 (635)                  | 149.61 | −3.61 | 6 | −0.54, −0.16 | *    |
| High                   | Low                        | 765 (452)                | 818 (474)                  | 242.11 | −1.06 | 8 | −0.22, 0.07  | 0.44 |
| High                   | Medium                     | 1035 (739)               | 1050 (564)                 | 112.07 | −0.99 | 4 | −0.37, 0.12  | 0.40 |
| High                   | High                       | 594 (316)                | 832 (592)                  | 231.02 | −3.37 | 8 | −0.41, −0.11 | **   |

\* $p < 0.05$ , \*\* $p < 0.01$  (FDR-corrected values). Comparisons were performed on the log-transformed total reading times. The means of raw reading times are provided for ease of interpretation, with the standard deviation in parenthesis.



who could likely extract the most meaning from the target multiword units. This pattern is not present in the group with lower L2 multiword scores (lighter columns), who were likely to extract less meaning and process the input at shallow level. Secondly, the groups reveal that the readers with lower L2 scores (again, lighter columns) were homogeneously fast in processing the more challenging L1-L2 *incongruent* items, and were even faster than readers with both high-L2 chunking ability and high-proficiency. Third, critically, high chunking ability learners only account for a small fraction of the faster total reading times. In fact, the data reveal that a considerable portion of the shortest reading times can be attributed to the readers with lower chunking ability, and among those, to the subset with lower L2 multiword knowledge.

One might also wonder how these modulators affected offline reading comprehension. Although all reading time analyses were performed on accurate trials only, the effect of L2 multiword proficiency and chunking ability on accuracy in offline responses was examined for each tercile-based subgroup. Interestingly, and in contrast with online reading times, the results of pairwise *t*-tests indicated that accuracy in offline comprehension was not significantly modulated by chunking ability (all  $p$ s > 0.80), but it was modulated by L2 proficiency instead. Lower-proficiency readers (mean accuracy: 85.3%, SD: 7.28) had significantly worse offline comprehension ( $p < 0.01$ ) than high-proficiency readers

(mean: 93.8%, SD: 5.12), and medium-proficiency readers (mean: 90.6%, SD: 5.52;  $p < 0.05$ ). This pattern further indicates a specialized role of chunking ability in predicting online processing efficiency, independent of other mechanisms that may affect the outcome of offline comprehension.

## Preamble Total Reading Times

Finally, one potential caveat affecting the generalizability of these results is that the focus of the analysis presented above was on N-that-V phrases and not on L2 reading more generally. To convincingly show the importance of chunking ability in reading, it would be necessary to demonstrate that this is a pervasive effect in sentence reading and not just on chunks (or modified chunks). The present data set allows to further investigate this question, by examining the overall reading times of the preamble sentences that preceded the critical sentences containing the multiword units<sup>9</sup>. Because the interest in this case is on the whole preamble sentence (and not on any specific manipulated region as in the

<sup>9</sup>I am grateful for the suggestion from one of the reviewers to analyze the preamble to address this point. Although the preamble sentences were not controlled or manipulated with any particular aspect of the L2 in mind, they were identical for all participants and therefore provide an additional opportunity to peruse the predictive power of individual Chunk Sensitivity during reading of the L2 more generally.

critical sentences), the total reading times of the preamble were used as the dependent measure (mean: 6,974 ms; SD: 3,391 ms).

If Chunk Sensitivity does indeed modulate reading efficiency more generally, the reading times of the preamble should reveal similar effects as the ones observed for the analysis of the critical multiword units. Recall that in the total reading times analyses of both the noun and the verb regions, the L2 chunking ability measure emerged as a significant modulator of total reading times; therefore, a similar effect in the preamble would provide compelling evidence of a general effect in reading. Based on the same data cleaning and analytical procedures described above, a mixed-effect model analysis was performed on the log-transformed total reading times of the preamble. The same variables as in the main analysis were included, with the exception of variables specifically related to the N-that-V multiword units (as these were not part of the preamble); the selected model is available in the **Supplementary Material**.

Importantly, L2 chunking ability was again revealed to be a significant predictor ( $\beta$ :  $-0.11$ , SE:  $0.05$ , CI:  $-0.20$ ,  $-0.02$ ,  $p < 0.05$ ). The model output is presented in **Table 6**. The results of this analysis provide further support for a general role of chunking ability in L2 reading.

## DISCUSSION

Based on previous evidence that L2 reading speed may not be a reliable indicator of ease of processing, this study investigated the influence that individual differences in L2 knowledge and processing efficiency bear on online reading. In the present study, online processing measures were acquired by recording L2 learners' eye movements during natural reading of Spanish sentences, which contained multiword units that were either congruent or non-congruent with their L1 (English). A recently developed measure of chunk sensitivity was employed as an index of processing efficiency in each of participants' two languages. Overall, the data confirmed the expected L1-L2 congruency effect during processing of the target verbs of multiword units, replicating previous findings of costlier processing of L1-L2 incongruent multiword units, relative to congruent items. In this sense, the present study elaborated on previous work by showing

that modified collocations, appearing as N-that-V relative noun phrases, yielded the same congruency effect reported in previous studies on V-N collocations.

However, the analysis revealed critical modulations of cross-linguistic influence at the individual level. At the group level, the results provided a clear replication of the L1 congruency effect reported for V-N phrases in which the verb is cross-linguistically incongruent (e.g., Wolter and Gyllstad, 2011, 2013). But the analysis of individual differences revealed that the effect was far from being homogeneously present in all readers. Rather, L1 and L2 chunk sensitivity measures, along with multiword-based proficiency, modulated processing costs between conditions as well as total reading speed. First, higher L1 chunk sensitivity was associated with faster reading times during early stages of processing, as captured by the measure of gaze duration. This finding was congruent with the only previously available study that examined the role of chunking ability in L2 processing (López-Beltrán et al., 2020). Secondly, the L2 measure of chunk sensitivity modulated performance in late stages of processing as indicated by total reading times. The analyses of total reading times of the critical verb region revealed an inverse U-shape effect, where the "poor chunkers" (low chunking-ability readers) who also had low L2 multiword-based proficiency showed some of the fastest reading times, but no significant differences between conditions, i.e., no congruency effect. This pattern indicated shallow processing, as can be gathered from the lower competence in the language and the less efficient processing of multiword chunks. On the other hand, individuals with high chunking ability and higher L2 multiword-based proficiency presented more efficient processing, with total faster reading times, as well as a significant difference between conditions. The intermediate values of the gamut presented a gradual slowdown in overall reading speed as knowledge of multiword units increased, but also a critical growth in between-condition differences as L2 chunking ability increased. Together, the results reveal joint contributions of knowledge of L2-specific multiword units and chunking ability, which modulate reading speed and processing efficiency. In what follows, the contribution of each factor is further discussed.

## Individual Chunking Ability Measures in the L1 and L2

One main goal of this study was to assess whether Chunk Sensitivity scores, either in the L1 or the L2, modulate reading times from L2 online processing measures (Research Question 1). The results indicated that individual scores of L1 Chunk Sensitivity were a significant predictor at early stages of L2 processing, as reflected by gaze duration. This is in line with the results reported by López-Beltrán et al. (2020), who found that L1 Chunk Sensitivity influenced self-paced reading times. It is relevant to note that gaze duration and self-paced reading both reflect the duration of a "first pass" in reading. In particular, because in self-paced paradigms the reader is not able to make regressive eye movements, this measure is *a priori* most comparable to gaze duration, which is computed based on the initial fixations on a region, before the eyes move on to

**TABLE 6 |** Summary of model output for reading times of the preamble.

| Fixed Effects                                   | $\beta$  | <i>SE</i> | <i>t</i>     | <i>p</i> |
|-------------------------------------------------|----------|-----------|--------------|----------|
| (Intercept)                                     | 8.74     | 0.06      | 150.63       | ***      |
| Eng. (L1) Chunk Sensitivity                     | 0.04     | 0.05      | 0.90         | 0.38     |
| Span. (L2) Chunk Sensitivity                    | −0.11    | 0.05      | −2.40        | *        |
| Random Effects                                  | Variance | <i>SD</i> | Correlations |          |
| Intercept   Participant                         | 0.10     | 0.32      |              |          |
| Trial Number   Participant                      | 0.00     | 0.01      | −0.49        |          |
| Intercept   Participant                         | 0.05     | 0.22      |              |          |
| Marginal $R^2$ : 0.05, Conditional $R^2$ : 0.69 |          |           |              |          |

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

the next word or move back to the left. The results reported here and in López-Beltrán and colleagues' study appear to converge in showing that chunk sensitivity measured in the *native* language predicts efficiency in early stages of *non-native* language processing. The association between chunking ability measured in the native language and processing of a language system acquired later in life would suggest that efficiency in early lexical access is dependent on domain-general retrieval mechanisms that are best captured by the L1 chunk sensitivity measure.

A novel contribution in the present study is the finding that L2 chunking ability was predictive of later stages of (second) language processing. That is, while L1 Chunk Sensitivity tended to modulate early access in gaze duration times, the L2 Chunk Sensitivity scores modulated total duration times, which are believed to primarily affect integration and incremental processing. In particular, while gaze duration is believed to index "initial" processing (e.g., early lexical access), total reading times are thought of as associated with later stages of processing, i.e., post-lexical access and aggregate effects of incremental processing and integration in context (Rayner, 1998). Because once bottom-up information is accessed (e.g., in lexical access in early stages of processing) the information needs to be structured and integrated within its context, knowledge of language-specific combinatory rules would be critical to guide the later stages of processing. Indeed, if chunking (i.e., binding of different elements during processing) is dependent on the projection of long-term memory representations that guide the scaffolded incremental processing of discourse segments, then language-specific chunk sensitivity should be critical in modulating later stages of processing. The significant effect of L2 chunk sensitivity in total durations is congruent with this view. Together, the effects of L1 and L2 chunk sensitivity measures depict a time-course of bilingual processing, in which domain-general and language-specific skills play a role at different steps of the incremental process of extracting meaning out of input.

Therefore, I suggest that the differences between chunk sensitivity measures in the native and non-native language are associated with discrete aspects involved in binding and integrating chunks of information in the input. How exactly, then, do the measures of chunking ability differ when measured in the L1 or in the L2 of language learners? Because the chunking ability task measures sensitivity to familiar "chunks" in a given language, the scores obtained reflect both language experience and domain-general chunking ability. The language experience component, i.e., familiarity with frequently co-occurring chunks in the language, necessarily relies on an individual's level of experience with a specific language. In this regard, an individual's chunk sensitivity score will not be the identical in each of the languages measured. Nonetheless, the measure is also believed to go beyond the static level of familiarity with given "chunks", and rather captures an individual's ability to recruit knowledge from prior experience to bind individual elements together and to build associations online. The multifaceted nature of the measure is reflected in the fact that L1 and L2 Chunk Sensitivity scores were not significantly correlated in the lower-intermediate learners tested in the present study. The only available benchmark against which these results can be compared are those reported

by López-Beltrán et al. (2020). In that study, the authors found a weak but significant correlation between the two versions of the task employed here ( $r = 0.37$ ,  $p = 0.01$ ). Although the exact same Chunk Sensitivity tasks were used in their study and the present one, an important difference is that the participants in the current study were recruited from basic Spanish courses (equivalent to the third and fourth semester), whereas López-Beltrán and colleagues tested students in upper-level and graduate Spanish courses. In accounting for the weak correlation between L1 and L2 chunk sensitivity, López-Beltrán and colleagues also speculated that L2 chunking ability may become a more reliable measure and more strongly correlated with L1 chunk sensitivity as L2 multiword knowledge increased. Their explanation is supported by the present dataset which indicates that, as they predicted, the association between chunk sensitivity in a L1 and L2 at lower proficiency levels is even weaker.

Finally, given the relative novelty of the Chunk Sensitivity measures, it is relevant to consider the validity of the tests, including aspects of predictive validity and construct validity. The extant evidence from L1 studies (McCauley and Christiansen, 2015; McCauley et al., 2017) and the first two L2 studies known to the author (i.e., this study and López-Beltrán et al.'s) has so far given robust support to the predictive validity of the tests. The dedicated analysis of subgroups (in the section "Follow-up analysis") provided further indication that chunking ability specifically predicted online processing, while offline comprehension was predicted by the L2 proficiency measure. Additionally, an important question relative to construct validity is whether the Chunk Sensitivity measures employed here are independent from other known cognitive measures, such as PSTM or working memory (WM). With regard to PSTM, correlational tests in the current dataset revealed no significant correlation between PSTM and chunking ability in either language (English:  $r = 0.05$ ,  $p = 0.76$ ; Spanish:  $r = 0.07$ ,  $p = 0.67$ ). Additional relevant confirmation of construct validity comes from López-Beltrán and colleagues, who investigated the contributions of chunking ability (measured in English and Spanish) and WM (measured through the Operation-Span task) to online processing, by including these measures in their analysis. Their results confirmed that chunking ability was a significant predictor even after WM was included in the model; importantly, WM did not emerge as a significant predictor of online processing. That is, in both the present study and López-Beltrán et al.'s work, chunking ability was a significant predictor of online reading times, whereas neither WM nor PSTM emerged as significant predictors of online processing. Taken together, the extant evidence provides support to the independent contribution of chunking ability, and the validity of the Chunk Sensitivity measures employed here beyond the well-established measures of PSTM and WM.

## Chunking Ability Is a Modulator of Cross-Linguistic Influence in L2 Processing

Two additional goals were to investigate in what ways individual knowledge of L2-specific units predicts differences in online

processing (RQ 2) and any potential interactions between L2 and chunking ability measures (RQ 3). The analysis took a different approach from previous studies that divided participants based on their overall reading speed (Hoover and Dwivedi, 1998; Roberts and Felser, 2011; Kaan et al., 2015). If both high- and low- efficient readers read fast, in a median split-based analysis, the individuals with the highest and lowest reading efficiency would be lumped together. In the absence of an adequate index of processing efficiency, previous studies found mixed evidence regarding the correlation of reading speed with proficiency measures (e.g., Roberts and Felser, 2011; Kaan et al., 2015). In the analysis reported here, total duration revealed a critical three-way interaction between L2 Chunk Sensitivity, L2 multiword knowledge and L1-L2 congruency in multiword units. Proficiency scores from L2 multiword knowledge –which tapped directly into the L2 aspect under consideration- were a significant predictor of gaze duration reading times.

A critical question was how chunking ability modulates cross-linguistic influences in processing (RQ 4). For incongruent multiword units, the results revealed the hypothesized inverse U-shaped pattern along the gradient of L2 chunk sensitivity, with faster reading times for individuals at both extremes of the chunking ability continuum (as illustrated in Figure 4). While, superficially, reading times may look similar in good and poor “chunkers,” the underlying causes are likely entirely different. That is, individuals with high chunking ability were believed to have more efficient online processing, in line with previous findings (McCauley and Christiansen, 2015; McCauley et al., 2017). On the other hand, poor chunking ability is associated with inefficient online computation of dependencies in the input, which may lead to faster (but shallow) reading. If this interpretation is correct, one would of course expect that shallow readers should be largely insensitive to the experimental manipulations and the properties of the input (i.e., cross-linguistic influence), while a more robust effect should be found in efficient processors. This prediction was confirmed by the analysis for readers at different levels of L2 multiword knowledge and chunking ability. Indeed, the results showed a lack of the well-attested L1-L2 congruency effect in the less advantaged readers. While individuals with poor L2 chunk sensitivity showed no significant effect of L1-L2 congruency, those with better chunking ability showed the effect. Remarkably, these patterns were not affected by L2 knowledge level. That is, the role of chunk sensitivity is highlighted by the finding that high chunking ability individuals showed the effect, regardless of L2 multiword knowledge, while the lower chunking ability readers showed a lack of congruency, even with higher L2 multiword knowledge. The results reported here provided evidence not only of shallow processing of information that requires L2-specific knowledge, but also showed that L2 chunk sensitivity can predict in what cases individual reading times will indicate shallow reading. In previous results reported by Kaan et al. (2015), the lack of the expected effects was interpreted as a clear indicator of shallow processing, but no correlates beyond the lack of differences between conditions were identified. The present results not only support this interpretation, but make a contribution by providing direct evidence of an

association between L2 chunking ability and the modulation of L2 processing costs in different conditions. In short, both good and poor “chunkers” may read similarly fast, but for entirely different reasons.

Importantly, the present findings highlight that the way in which cross-linguistic influence emerges and is modulated by chunking ability may be non-linear. Clearly, the emergence of cross-linguistic differences requires a certain level of depth in processing; during shallow reading, in which the reader does not engage with L2-specific features in the input, expected effects of cross-linguistic influence may not be revealed. On the other end, with increasing chunking ability, the expected cross-linguistic influence is reduced as a consequence of more efficient processing. Future work will be needed to confirm this pattern by examining whether and how chunking ability modulates cross-linguistic differences, while also carefully controlling for individuals’ command of the specific L2 feature under examination. Testing how chunking ability impacts processing of phonological, morphosyntactic or lexical features will be important in further characterizing the scope of this modulator.

## How Generalizable Are Group-Based Findings? Implications for Research on L2 Processing

Based on the group averages, which replicated the well-attested L1-L2 congruency effect (e.g., Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013; Carrol et al., 2016; Wolter and Yamashita, 2018), one could examine group-based grand averages, find the results satisfactory and in no need for further investigation. However, this would tell us only part of the story – and an imperfect one – about L2 processing and the impact of cross-linguistic influence. The results reported here evince that, without taking into consideration and appropriately accounting for individual differences, the congruency effect of the grand average is driven by about half of the learners in this sample. This would be, by all means, an inaccurate result and an inappropriate depiction of L2 processing in the tested population.

Given the critical importance of developing theories that are generalizable (i.e., extrapolated from a representative sample to the “real” population), reaching an accurate understanding of a given sample on which to ground our conclusions is imperative. The findings reported here provide evidence that group-based average reading times may provide a poor depiction of the individuals therein. As in previous L2 research, the results revealed substantial variability within the sample tested. By identifying chunking ability as a modulator of reading, the analysis identified systematic influences on reading patterns within the sample. The variability uncovered by the analyses highlights two striking points. First, the large amount of variability, with about half of the participants showing between-group differences but with no significant effect in about half of the sample, underscores the coarseness of the group-based measure. Secondly, given the independent contribution of chunking ability, the fact that this and previous studies have reliably replicated a congruency effect is in itself remarkable. The important point is that strong effects may be replicable even if

they are far less frequent in the general population. To illustrate, based on the present findings, one could expect the congruency effect to be reliably found at the group level (given a large enough sample) but, at the level of the individual learner, it could be expected to be found in about half of the cases. Therefore, it is not only important to control for potential idiosyncracies present within a given sample, but also to understand and account for the existence of variety within the group<sup>10</sup>.

In summary, the findings reported here call for a renewed attempt in processing studies to characterize individual variability within groups of participants. Provided sufficiently large sample sizes, the approach presented here integrating measures of chunking ability is an invitation to reanalyze datasets and to, potentially, better characterize the performance of subgroups of individuals within the sample. While efforts to explore the role of chunking ability are just beginning, the extant evidence already suggests that its influence affects L2 processing of aspects as varied as cross-language congruency in multiword units (in the present results) and morphosyntactic expression of gender (López-Beltrán et al., 2020). Future studies should investigate how the clear association between multiword chunking ability and processing explored here will hold when other linguistic aspects are examined, e.g., in L2 syntax or morphosyntax; future work should also pursue conceptual replications using different sets of multiword units and including also comparisons to L1 speakers. Large-scale studies as well as work in other languages will be instrumental to further validate chunking-based measures; for instance, interesting questions arise regarding chunking in languages with different writing systems. The data available make a case for the consideration of chunking ability as a critical modulator of individual differences in L2 reading, and in language processing more generally.

## CONCLUSION

This article addresses a gap in previous studies which reported signs of shallow processing during L2 reading, but found that neither language proficiency nor reading speed alone were reliable predictors of online reading performance. Instead, the results in this study identified chunking ability as a critical modulator of online L2 processing and of cross-linguistic influences in particular. Crucially, the data highlight the fact that only a fraction of the fastest reading speeds can be attributed to high processing efficiency (i.e., chunk sensitivity) or to higher L2 experience, while many other fast reading times are directly associated with poor chunking ability. Altogether, the findings suggest that incorporating measures of chunking ability in the L1 and L2 might add a fundamental dimension to account for processing effects, not only at the level of the individual but

<sup>10</sup> Additional factors may also influence individuals differently, e.g., it may also be that the reversed order of collocations in N-that-V phrases posed greater demands on processing than canonical V-N phrases would, perhaps affecting particularly individuals with low chunking ability. Relatedly, individuals may show different sensitivity to other stimulus features (e.g., incongruent collocations tend to be less transparent; Yamashita, 2018). More research will be needed to further investigate how features such as the degree of familiarity or transparency may interact with chunking ability.

also in the aggregate – i.e., at the group level. Future replication studies, including potential work on more typologically distant languages, will be required to confirm the role of L1 and L2 chunking ability. The present results are particularly promising given that the modulators of L2 reading efficiency have proven to be an elusive target in previous work. Analyses that miss this modulator may be unable to capture processing efficiency as a true source of variability. In conclusion, the findings have potentially far-reaching implications for the interpretation of previous and future results derived from online reading times, and invite future work to explore the contribution of individual chunking ability by factoring it into investigations of L2 online processing.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board, Penn State University. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## FUNDING

This work was supported by the NSF Doctoral Dissertation grant BCS-1844188 and a *Language Learning* Doctoral Dissertation Award.

## ACKNOWLEDGMENTS

The author would like to thank Katrina Connell, Mary Henderson, Anna Northcott, Zeal Patel, Enissa Ramos, Laura Rodrigo, and Tanya Wang for their help with data collection. Thanks to Christian Navarro-Torres, as well as to Tamar Degani and the two reviewers of this article, for their very helpful comments on earlier versions of this manuscript. Special thanks to Paola E. Dussias for her guidance and invaluable advice at various stages of the project. Any errors are my own.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.607621/full#supplementary-material>

## REFERENCES

- Alderson, J. C. (1984). "Reading in a foreign language: a reading problem or a language problem?," in *Reading in a Foreign Language*, eds J. C. Alderson and A. H. Urquhart (London: Longman), 1–24.
- Alderson, J. C. (1993). "The relationship between grammar and reading in an English for academic purposes test battery," in *A New Decade of Language Testing Research: Selected Papers from the 1990 Language Testing Research Colloquium: Dedicated in Memory of Michael Canale*, eds D. Douglas and C. Chapelle (Alexandria, VA: Teachers of English to Speakers of Other Languages), 203–219.
- Baddeley, A., Papagno, C., and Vallar, G. (1988). When long-term learning depends on short-term storage. *J. Mem. Lang.* 27, 586–595. doi: 10.1016/0749-596X(88)90028-9
- Baddeley, A. D. (1979). *Working Memory and Reading*. In *Processing of Visible Language*. Boston, MA: Springer, 355–370.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.
- Bowey, J. A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *J. Exp. Child Psychol.* 41, 282–299. doi: 10.1016/0022-0965(86)90041-x
- Broughton, S. H., Sinatra, G. M., and Reynolds, R. E. (2010). The nature of the refutation text effect: an investigation of attention allocation. *J. Educ. Res.* 103, 407–423. doi: 10.1080/00220670903383101
- Budiu, R., Royer, C., and Pirolli, P. L. (2007). "Modeling information scent: a comparison of LSA, PMI and GLSA similarity measures on common tests and corpora," in *Proceedings of the 8th Annual Conference of the Recherche d'Information Assistée par Ordinateur (RIA/O)*, (Pittsburgh, PA: Centre des Hautes Études Internationales d'Informatique Documentaire).
- Bullinaria, J. A., and Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: a computational study. *Behav. Res. Methods* 39, 510–526. doi: 10.3758/bf03193020
- Carrol, G., and Conklin, K. (2017). Cross language lexical priming extends to formulaic units: evidence from eye-tracking suggests that this idea "has legs." *Biling. Lang. Cogn.* 20, 299–317. doi: 10.1017/S1366728915000103
- Carrol, G., Conklin, K., and Gyllstad, H. (2016). Found in translation: the influence of the L1 on the reading of idioms in a L2. *Stud. Second Lang. Acquis.* 38, 403–433. doi: 10.1017/S0272263115000492
- Christiansen, M. H., and Chater, N. (2016). The Now-or-Never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* 39, 1–72. doi: 10.1017/S0140525X1500031X
- Cieri, C., David, M., and Kevin, W. (2004). "The fisher corpus: a resource for the next generations of speech-to-text," in *Proceedings of the 4th International Conference on Language Resources and Evaluation*, (Lisbon). doi: 10.1007/978-94-017-1183-8\_1
- Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/s0142716406060024
- Coughlin, C. E., and Tremblay, A. (2012). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Appl. Psycholinguist.* 34, 615–646. doi: 10.1017/S0142716411000890
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Mem. Lang.* 19, 450–466.
- Davies, M. (2016). *Corpus del Español: Two billion words, 21 Countries*. Available online at <http://www.corpusdelespanol.org/web-dial/> (accessed October 15, 2018).
- Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annu. Rev. Appl. Linguist.* 30, 149–166. doi: 10.1017/s026719051000005x
- Elliott, L. L. (1962). Backward and forward masking of probe tones of different frequencies. *J. Acoust. Soc. Am.* 34, 1116–1117. doi: 10.1121/1.1918254
- Erçetin, G., and Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: implications for second-language reading comprehension. *Appl. Psycholinguist.* 34, 727–753. doi: 10.1017/s0142716411000932
- Felser, C., Roberts, L., Marinis, T., and Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Appl. Psycholinguist.* 24, 453–489. doi: 10.1017/s0142716403000237
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences Cognitive Psychology. *Cogn. Psychol.* 47, 164–203. doi: 10.1016/S0010-0285(03)00005-7
- Ferreira, F., Christianson, K., and Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of sentence processing and reanalysis. *J. Psycholinguist. Res.* 30, 3–20.
- Ferreira, F., and Patson, N. D. (2007). The?Good Enough? approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83. doi: 10.1111/j.1749-818x.2007.00007.x
- Foucart, A., and Frenck-Mestre, C. (2011). Grammatical gender processing in L2: electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism* 14, 379–399. doi: 10.1017/S136672891000012X
- Fox, J., Friendly, G. G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., et al. (2007). *The Car Package*. Boston, MA: R Foundation for Statistical Computing.
- Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Lang. Res.* 21, 175–198. doi: 10.1191/0267658305sr2570a
- Gathercole, S. E., Pickering, S. J., Hall, M., and Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Q. J. Exp. Psychol. Section A* 54, 1–30. doi: 10.1080/02724980042000002
- Green, P., and MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498. doi: 10.1111/2041-210X.12504
- Harrington, M., and Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Stud. Second Lang. Acquisit.* 14, 25–38. doi: 10.1017/s0272263100010457
- Havik, E., Roberts, L., Van Hout, R., Schreuder, R., and Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Lang. Learn* 59, 73–112. doi: 10.1111/j.1467-9922.2009.00501.x
- Hoover, M., and Dwivedi, V. (1998). Syntactic processing by skilled bilinguals. *Lang. Learn.* 48, 1–29. doi: 10.1111/1467-9922.00031
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Lang. Res.* 22, 369–397. doi: 10.1191/0267658306sr2720a
- Hopp, H. (2010). Ultimate attainment in L2 inflection: performance similarities between non-native and native speakers. *Lingua* 120, 901–931. doi: 10.1016/j.lingua.2009.06.004
- Hopp, H. (2013). Grammatical gender in adult L2 acquisition: relations between lexical and syntactic variability. *Second Lang. Res.* 29, 33–56. doi: 10.1177/0267658312461803
- Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Lang. Acquisit.* 21, 250–278. doi: 10.1080/10489223.2014.892943
- Hopp, H. (2015). The timing of lexical and syntactic processes in L2 sentence comprehension. *Appl. Psycholinguist.* 39, 1–28.
- Hsu, N. S., and Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychol. Sci.* 27, 572–582. doi: 10.1177/0956797615625223
- Izura, C., Cuertos, F., and Brysbaert, M. (2014). Lextale-Esp: a test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica* 35, 49–66.
- Jeon, E. H., and Yamashita, J. (2014). L2 reading comprehension and its correlates: a meta-analysis. *Lang. Learn.* 64, 160–212. doi: 10.1111/lang.12034
- Jiang, N. (2015). "Six decades of research on bilingual presenation," in *The Cambridge Handbook of Bilingual Processing*, ed. J. W. Schwieter (Cambridge: Cambridge University Press), 29–84. doi: 10.1017/cbo9781107447257.002
- Juffs, A. (2004). Representation, processing and working memory in a second language. *Trans. Philol. Soc.* 102, 199–225. doi: 10.1111/j.0079-1636.2004.00135.x
- Just, M. A., and Carpenter, P. A. (1980). "A theory of reading: from eye fixations to comprehension," in *Psychological Review*, ed. K. J. Holyoak (Washington, D.C: American Psychological Association), 87.

- Kaan, E., Ballantyne, J. C., and Wijnen, F. (2015). Effects of reading speed on second-language sentence processing. *Appl. Psycholinguist.* 36, 799–830. doi: 10.1017/S0142716413000519
- Karimi, H., and Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Q. J. Exp. Psychol.* 69, 1013–1040. doi: 10.1080/17470218.2015.1053951
- Keating, G. D., and Jegerski, J. (2015). Experimental designs in sentence processing research: a methodological review and user's guide. *Stud. Second Lang. Acquisit.* 37, 1–32. doi: 10.1017/s0272263114000187
- Kieffer, M. J., and Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Read. Writ. Interdiscip. J.* 21, 783–804. doi: 10.1007/s11145-007-9092-8
- Kieffer, M. J., and Lesaux, N. K. (2012a). Direct and indirect roles of morphological awareness in the English reading comprehension of native Spanish, Filipino, Vietnamese, and English speakers. *Lang. Learn.* 62, 1170–1204. doi: 10.1111/j.1467-9922.2012.00722.x
- Kieffer, M. J., and Lesaux, N. K. (2012b). Knowledge of words, knowledge about words: dimensions of vocabulary in first and second language learners in sixth grade. *Read. Writ.* 25, 347–373. doi: 10.1007/s11145-010-9272-9
- Koda, K. (2005). *Insights into Second Language Reading: A Cross-Linguistic Approach*. New York, NY: Cambridge University Press.
- Kotz, S. A., and Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *J. Neurolinguist.* 17, 215–235. doi: 10.1016/S0911-6044(03)00058-7
- Kyriacou, M., Conklin, K., and Thompson, D. (2020). Passivizability of idioms: has the wrong tree been barked up? *Lang. Speech* 63, 404–435. doi: 10.1177/0023830919847691
- Laufer, B., and Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Appl. Linguist.* 29, 694–716. doi: 10.1093/applin/amn018
- Ley, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013
- Long, J. A. (2019). *Interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions. R package version 1.1.0*. Available online at: <https://cran.r-project.org/package=interactions> (accessed July 20, 2020).
- López-Beltrán, P., Pulido, M. F., Dussias, P. E., and Christiansen, M. (2020). *Individual Differences in Chunking Ability Predict Native-Like Second Language Processing: Evidence from Spanish Foreigner Agreement*. Submitted for Publication. Available online at: [https://osf.io/vx5g8/?view\\_only=48ba0e2fdb14ad4b182ac9f90cfea13](https://osf.io/vx5g8/?view_only=48ba0e2fdb14ad4b182ac9f90cfea13) (accessed July 20, 2020).
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marsden, E., Thompson, S., and Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Appl. Psycholinguist.* 39, 861–904. doi: 10.1017/s0142716418000036
- Martin, K. I., and Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Stud. Second Lang. Acquisit.* 34, 379–413. doi: 10.1017/S0272263112000125
- McCauley, S. M., and Christiansen, M. H. (2015). “Individual differences in chunking ability predict on-line sentence processing individual differences in chunking ability predict on-line sentence processing,” in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, (Pasadena, CA), 1553–1558.
- McCauley, S. M., Isbilen, E. S., and Christiansen, M. H. (2017). “Chunking ability shapes sentence processing at multiple levels of abstraction,” in *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, eds G. Gunzelmann, A. Howes, T. Tenbrink, and E. J. Davelaar (Austin, TX: Cognitive Science Society), 2681–2686. Available online at: <http://cnl.psych.cornell.edu/pubs/2017-mic-cogsci.pdf>
- Metzner, P., von der Malsburg, T., Vasishth, S., and Rösler, F. (2017). The importance of reading naturally: evidence from combined recordings of eye movements and electric brain potentials. *Cogn. Sci.* 41, 1232–1263. doi: 10.1111/cogs.12384
- Miller, B. W. (2015). Educational psychologist using reading times and eye-movements to measure cognitive engagement. *Educ. Psychol.* 50, 31–42. doi: 10.1080/00461520.2015.1004068
- Miller, G. A., and Taylor, W. G. (1948). The perception of repeated bursts of noise. *J. Acoust. Soc. Am.* 20, 171–182. doi: 10.1121/1.1906360
- Navarro-Torres, C. A., Garcia, D. L., Chidambaram, V., and Kroll, J. F. (2019). Cognitive control facilitates attentional disengagement during second language comprehension. *Brain Sci.* 9, 19–21. doi: 10.3390/brainsci9050095
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English. *Appl. Linguist.* 24, 223–242. doi: 10.1093/applin/24.2.223
- Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., and Bunting, M. F. (2013). Clearing the garden-path: improving sentence processing through cognitive control training. *Lang. Cogn. Neurosci.* 29, 186–217. doi: 10.1080/01690965.2012.758297
- Pashler, H. (1988). Familiarity and visual change detection. *Percept. Psychophys.* 44, 369–378. doi: 10.3758/bf03210419
- Payne, T. W., Kalibateva, Z., and Jungers, M. K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension? *Learn. Individ. Differ.* 19, 119–123. doi: 10.1016/j.lindif.2008.05.003
- Pulido, M. F. (2020). Native language inhibition predicts more successful second language learning: evidence of two ERP pathways during learning. *Neuropsychologia* doi: 10.1016/j.neuropsychologia.2020.107732 [Epub ahead of print].
- Pulido, M. F., and Dussias, P. E. (2020). Desirable difficulties while learning collocations in a second language: conditions that induce L1 interference improve learning. *Biling. Lang. Cogn.* 23, 652–667. doi: 10.1017/S1366728919000622
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Recchia, G., and Jones, M. N. (2009). More data trumps smarter algorithms: comparing pointwise mutual information with latent semantic analysis. *Behav. Res. Methods* 41, 647–656. doi: 10.3758/BRM.41.3.647
- Remez, R. E., Ferro, D. F., Dubowski, K. R., Meer, J., Broder, R. S., and Davids, M. L. (2010). Is desynchrony tolerance adaptable in the perceptual organization of speech? *Atten. Percept. Psychophys.* 72, 2054–2058. doi: 10.3758/bf03196682
- Reppen, R., Ide, N., and Suderman, K. (2005). *American National Corpus. Philadelphia: Linguistic Data Consortium*. Available online at: <http://americannationalcorpus.org/>
- Roberts, L. (2012). Individual differences in second language sentence processing. *Lang. Learn.* 62, 172–188. doi: 10.1111/j.1467-9922.2012.00711.x
- Roberts, L., and Felsner, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Appl. Psycholinguist.* 32, 299–331. doi: 10.1017/S0142716410000421
- Sagarra, N., and Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua* 120, 2022–2039. doi: 10.1016/j.lingua.2010.02.004
- Shiotsu, T. (2010). *Components of L2 Reading: Linguistic and Processing Factors in The Reading Test Performances of Japanese EFL Learners*. New York, NY: Cambridge University Press.
- Siegel, L. S. (1994). Working memory and reading: a life-span perspective. *Int. J. Behav. Dev.* 17, 109–124. doi: 10.1177/016502549401700107
- Siyanova-Chanturia, A., Conklin, K., and Schmitt, N. (2011a). Adding more fuel to the fire: an eye-tracking study of idiom processing by native and non-native speakers. *Second Lang. Res.* 27, 251–272. doi: 10.1177/0267658310382068
- Siyanova-Chanturia, A., Conklin, K., and van Heuven, W. J. B. (2011b). Seeing a Phrase “Time and Again” Matters: the role of phrasal frequency in the processing of multiword sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 776–784. doi: 10.1037/a0022531
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* 80, 99–103. doi: 10.1207/S15327752JPA8001\_18
- Tavakol, M., and Dennick, R. (2011). Making sense of Cronbach's alpha. *Int. J. Med. Educ.* 2, 53–55. doi: 10.5116/ijme.4dfb.8dfd
- Tokowicz, N., and MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Stud. Second Lang. Acquisit.* 27, 173–204.

- Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *J. Exp. Psychol. Learn. Mem. Cogn.* 42:1632. doi: 10.1037/xlm0000259
- Vilkaite, L., and Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Appl. Linguist.* 40, 329–354. doi: 10.1093/applin/amx030
- Wolter, B., and Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Appl. Linguist.* 32, 430–449. doi: 10.1093/applin/amr011
- Wolter, B., and Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Stud. Second Lang. Acquisit.* 35, 451–482. doi: 10.1017/S0272263113000107
- Wolter, B., and Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: what accounts for L2 performance? *Stud. Second Lang. Acquisit.* 40, 395–416. doi: 10.1017/S0272263117000237
- Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *J. Second Lang. Stud.* 1, 60–78. doi: 10.1075/jsls.17024.yam
- Yamashita, J., and Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Q.* 44, 647–668. doi: 10.5054/tq.2010.235998

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Pulido. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Costs and Benefits of Native Language Similarity for Non-native Word Learning

Viorica Marian<sup>1\*</sup>, James Bartolotti<sup>2</sup>, Aimee van den Berg<sup>1</sup> and Sayuri Hayakawa<sup>1</sup>

<sup>1</sup> Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States, <sup>2</sup> Life Span Institute, University of Kansas, Lawrence, KS, United States

## OPEN ACCESS

### Edited by:

Tamar Degani,  
University of Haifa, Israel

### Reviewed by:

Natasha Tokowicz,  
University of Pittsburgh, United States  
Agnieszka Otwinowska-Kasztelanic,  
University of Warsaw, Poland

### \*Correspondence:

Viorica Marian  
v-marian@northwestern.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 January 2021

**Accepted:** 29 April 2021

**Published:** 28 May 2021

### Citation:

Marian V, Bartolotti J, van den Berg A  
and Hayakawa S (2021) Costs and  
Benefits of Native Language Similarity  
for Non-native Word Learning.  
Front. Psychol. 12:651506.  
doi: 10.3389/fpsyg.2021.651506

The present study examined the costs and benefits of native language similarity for non-native vocabulary learning. Because learning a second language (L2) is difficult, many learners start with easy words that look like their native language (L1) to jumpstart their vocabulary. However, this approach may not be the most effective strategy in the long-term, compared to introducing difficult L2 vocabulary early on. We examined how L1 orthographic typicality affects pattern learning of novel vocabulary by teaching English monolinguals either Englishlike or Non-Englishlike pseudowords that contained repeated orthographic patterns. We found that overall, the first words that individuals learned during initial acquisition influenced which words they acquired later. Specifically, learning a new word in one session made it easier to acquire an orthographically similar word in the next session. Similarity among non-native words interacted with native language similarity, so that words that looked more like English were easier to learn at first, but they were less effective at influencing later word learning. This demonstrates that although native language similarity has a beneficial effect early on, it may reduce learners' ability to benefit from non-native word patterns during continued acquisition. This surprising finding demonstrates that making learning easier may not be the most effective long-term strategy. Learning difficult vocabulary teaches the learner what makes non-native words unique, and this general wordform knowledge may be more valuable than the words themselves. We conclude that native language similarity modulates new vocabulary acquisition and that difficulties during learning are not always to be avoided, as additional effort early on can pay later dividends.

**Keywords:** language learning, cross-language similarity, second language, language acquisition, vocabulary learning

## INTRODUCTION

Children often excel at learning new languages—consider international adoptees who rapidly acquire their “second first language” (Roberts et al., 2005)—whereas for adults, learning a second language (L2) has traditionally been thought to be a more difficult task (Liskin-Gasparro, 1982). There is now substantial evidence that, for children and adults alike, the ability to successfully learn a second language can be moderated by complex interactions between contextual, sociocultural, cognitive, and affective variables (see Dixon et al., 2012 and Ortega, 2013 for reviews), as well as characteristics of the first (L1) and second (L2) languages. Oftentimes, learners can take advantage of similarities between the L1 and L2 by relying on existing skills and knowledge to learn the

new language (i.e., cross-linguistic transfer or cross-linguistic influence; Ringbom, 2007; Jarvis and Pavlenko, 2008). Other times, transfer from the L1 can inhibit learning, such as when L1 knowledge is inappropriately applied (Laufer, 1988; Eckman, 2004) or when it interferes with the acquisition of L2-specific representations (Flege, 1987; Goldrick et al., 2014). Seeing language acquisition as an incremental process (that successively builds on previously-learned information), the costs and benefits of cross-linguistic influences at early stages of learning could have cascading consequences for later acquisition. In the current study, we examine the developmental trajectory of cross-linguistic influences on novel word learning and the role of orthographic similarity to previously-learned native and non-native words.

## Effects of Cross-Linguistic Transfer on New Word Learning

Native language similarity has long been known to be a powerful resource for language learning. In many cases, the ease and efficiency of language acquisition can be modulated by existing language knowledge and the actual (Ringbom, 2007) or perceived (Kellerman, 1978; Odlin, 1989) formal similarities between languages. Similarities between languages and cross-linguistic transfer can be found at multiple levels of representation, such as phonology (e.g., Melby-Lervåg and Lervåg, 2011; Wrembel, 2011), orthography (e.g., De Groot and Keijzer, 2000; Ellis, 2008), and morphology (e.g., Hancin-Bhatt and Nagy, 1994; Ecker, 2015). During vocabulary learning, cognates, which overlap across languages in both orthographic form and meaning, are often more readily acquired than non-cognates (Lotto and de Groot, 1998; De Groot and Keijzer, 2000). In addition to allowing learners to draw from existing knowledge, L1-L2 similarity may facilitate integration of novel wordforms into the existing lexico-semantic network (Shirai, 1992; MacWhinney, 1997; Comesaña et al., 2009; De Groot, 2011), which can result in more robust encoding during early stages of acquisition (Ellis and Beaton, 1993), as well as more fluid retrieval at higher levels of proficiency (e.g., Comesaña et al., 2012).

SLA studies conducted in more naturalistic contexts (e.g., classrooms) have also found advantages for learning cognates (e.g., Cunningham and Graham, 2000; Tonzar et al., 2009; Vidal, 2011; Otwinowska and Szewczyk, 2019; Puimège and Peters, 2019), but have produced more mixed results (e.g., Rogers et al., 2015; Otwinowska et al., 2020; see Otwinowska, 2015 for review). Some evidence suggests that cognate facilitation may be contingent on formal training in cognate recognition (Tréville, 1996; Dressler et al., 2011), suggesting that learners may not always be aware of formal similarities. Indeed, using contrastive analysis to highlight similarities and differences between the L1 and L2 can be highly effective (Laufer and Girsai, 2008; Lin, 2015; Helms-Park and Perhan, 2016), and students tend to respond positively to this type of language instruction (Brooks-Lewis, 2009). Additionally, the trade-off between ecological validity and control over stimulus characteristics (e.g., word frequency, orthographic overlap) may also contribute to the more tenuous cognate effects found in SLA studies compared to laboratory

experiments (but see Otwinowska and Szewczyk, 2019 for an exception and Otwinowska, 2015 for discussion).

Psycholinguistic studies using carefully controlled real or artificial word stimuli have revealed that the cognate advantage increases with the degree of orthographic overlap (De Groot, 2011; Comesaña et al., 2015), and that even without semantic overlap, vocabulary acquisition can be facilitated for novel words that are orthographically (De Groot, 2011; Bartolotti and Marian, 2014, 2017; Marecka et al., 2021) or phonologically (Ellis and Beaton, 1993; Service and Craik, 1993; Roodenrys and Hinton, 2002; Storkel and Maekawa, 2005; Storkel et al., 2006) similar to L1 words. For instance, Meade et al. (2018) observed that pseudowords with a higher number of L1 orthographic neighbors were produced more accurately than low-density words. Like cognates, the advantage for words that resemble L1 wordforms could result from more effective use of, and integration with, existing lexical and semantic knowledge. This can be done explicitly through association-based strategies (e.g., mnemonic methods; Atkinson and Raugh, 1975; Meara, 1980; Paivio and Desrochers, 1981; see Hulstijn, 1997 and Nation, 1982 for reviews) or implicitly through the co-activation of orthographically or phonologically similar L1 words and associated meanings (Holcomb et al., 2002; Van Hell and Tanner, 2012).

Words with familiar orthographic or phonological features may additionally benefit word learning by allowing learners to exploit knowledge of sublexical regularities in the L1. For instance, Bartolotti and Marian (2014) found that vocabulary acquisition is facilitated when pseudowords are designed to reuse native language letter patterns (i.e., higher bigram probabilities; see also Bartolotti and Marian, 2017). Phonological overlap between the L2 and L1 can additionally facilitate learning by increasing pronounceability (Ellis and Beaton, 1993; Service and Craik, 1993), which could enable learners to rely on phonological knowledge stored in long-term memory (Cheung, 1996; Storkel and Maekawa, 2005; De Groot, 2011) and the mental rehearsal of novel phonological forms (Papagno et al., 1991; Ellis and Sinclair, 1996).

Despite potential advantages, lexical and sublexical similarities between languages can also introduce costs when they are over-applied or block acquisition of new features during learning. For example, a German learner of English may say, “I need a loffel for my soup,” under the mistaken belief that the German word *Löffel* (meaning spoon) is an English cognate (Eckman, 2004). This type of confusion can be especially likely when a novel word overlaps with a known word in some respects, but not others (i.e., “deceptive transparency;” Laufer, 1988), such as when an L2 word overlaps in form but not meaning with an L1 word (i.e., false friends; e.g., the German word *Rat*, which means “advice”). A different problem occurs when similarity to the L1 interferes with acquisition of L2-specific features or regularities, as seen with spoken accents. Sounds in the L2 that are similar to an existing L1 sound are actually *more* difficult to pronounce accurately than completely new sounds (Flege, 1987). Even speakers who have mastered L2 phonology still pronounce cognate words with more of an accent than non-cognates, due to cognates’ high L1 similarity (Amengual, 2012,

2016; Goldrick et al., 2014; consistent with Anderson's (1983) *transfer to somewhere* principle).

Kaushanskaya and Marian (2009) found that, compared to when phonologically atypical pseudowords were presented alone, learning was impaired when words were presented bimodally with *typical* L1 orthographic forms. Such effects could be explained by the increased activation of L1 representations in response to familiar wordforms, which could compete with the more recently acquired L2 representation. For instance, psycholinguistic studies have demonstrated that visual word recognition can be inhibited by orthographically-related primes (both within and across languages; e.g., Bijeljac-Babic et al., 1997) that can compete for selection. Cross-language lexical activation can be reduced, however, by priming different script bilinguals (e.g., Hindi-English) with a particular writing system (Dubey et al., 2018), as well as by priming same script bilinguals with language-specific sublexical cues (e.g., bigrams) that are uncommon or orthotactically illegal in one language but not the other (Casaponsa et al., 2020). Similarly, language discrimination is facilitated by orthographic markers that signal language membership (Casaponsa et al., 2014; Oganian et al., 2015). These patterns of facilitation and interference indicate a high degree of cross-linguistic interactivity within the language system, which can play a significant role in vocabulary acquisition and be amplified by similarities at the lexical and sublexical level of processing.

## Effects of Within-Language Transfer on New Word Learning

Similarity to the L1 can yield significant benefits during early stages of word learning by encouraging cross-linguistic transfer. The recognition and use of L2-specific patterns, however, is key to long-term success in developing L2 vocabulary. Adults who had completed 1 year of university-level Spanish courses were able to learn new words with a large number of Spanish neighbors (i.e., words that differed from many Spanish words by only a single phoneme) at a higher rate than words with a low number of Spanish neighbors (Stamer and Vitevitch, 2012). This ability to learn words with more L2 neighbors provides evidence that similarities *within* an L2 benefit learning. The application of within-language knowledge for novel word learning can additionally vary as a function of individual differences such as L2 proficiency (e.g., Horst et al., 1998; Zahar et al., 2001; Pulido, 2003; Tekmen and Daloglu, 2006; Ma et al., 2015; Otwinowska and Szweczyk, 2019). For instance, Ma et al. (2015) observed that L2 proficiency was positively associated with learners' ability to learn the meanings of novel pseudowords embedded in sentences. Studies employing incidental learning paradigms have similarly observed that higher L2 proficiency and larger L2 vocabulary sizes facilitate novel vocabulary acquisition during reading (e.g., Horst et al., 1998; Tekmen and Daloglu, 2006). Such findings suggest that as proficiency in the L2 increases, so too does the strength of within-L2 facilitation, creating a positive feedback loop where L2 word learning becomes easier as L2 vocabulary size increases. Proficiency can also modulate the contribution of other domain-general cognitive abilities

(Cheung, 1996; Gathercole and Masoura, 2005). For instance, Cheung (1996) found that greater short-term memory capacity was associated with better L2 vocabulary learning for individuals with low, but not high L2 proficiency (but see Majerus et al., 2008 who found independent effects of STM and L2 phonological knowledge). Bartolotti et al. (2011) observed that inhibitory control and bilingual experience independently predicted how well learners were able to extract statistical regularities of word boundaries in an artificial Morse Code language after listening to another language with conflicting patterns. Bilingual experience facilitated word learning when interference was low, whereas inhibitory control predicted performance when interference was high (see also Wang and Saffran, 2014, who observed a bilingual advantage for detecting regularities in an artificial tonal language). The ability to extract and apply regularities within the L2 can therefore vary depending on both individual differences in cognitive and linguistic abilities, as well as characteristics of the learning task.

Though gains are likely to compound with increased L2 experience, the beneficial effect of within-L2 similarity applies even at the earliest stages of acquisition (McLaughlin et al., 2004; Bartolotti and Marian, 2017). After only 14 h of classroom study, novice L2 learners' neural responses indicated familiarity with words they had seen before, even though behaviorally they only identified words at chance performance (McLaughlin et al., 2004; Osterhout et al., 2006). After only one session of training in an artificial language, learners demonstrate that they have learned letters' relative frequencies in the language, and can use this information to fill gaps in their knowledge of the new language (Bartolotti and Marian, 2017). Other statistical regularities governing word boundaries can be learned from continuous speech after as little as 20 min of exposure (Saffran et al., 1999; Karuza et al., 2013), and this knowledge of word boundaries can directly influence subsequent vocabulary acquisition (Mirman et al., 2008). Together, these findings demonstrate that learners are able to extract L2 regularities based on even brief amounts of exposure, which can then be used to support further learning.

## Effects of Cross-Linguistic Influence on Within-Language Transfer

While there has been substantial research investigating the independent effects of between- and within-language transfer on vocabulary acquisition, relatively less is known about their potential interactions—specifically, whether native language orthographic similarity modulates transfer between non-native words during subsequent learning. A significant body of research has shown that the strategies individuals use to process words within the L1 is influenced by their orthographic system (e.g., Hakuta, 1982; MacWhinney and Bates, 1989), and that these same processes may be used to decode words in an L2 (e.g., Koda, 1998; Mori, 1998; Hamada and Koda, 2008). As a result, sensitivity to L2-specific orthotactics can vary as a function of similarity between L1 and L2 orthographic systems. For instance, Koda (1998) observed that ESL learners with L1 Korean (who utilize a syllable-based writing system, *hangul*) were more sensitive to English orthotactic violations (i.e.,

illegal letter sequences) than L1 Chinese speakers (who utilize a morpheme-based logographic writing system). The author conjectures that the Korean speakers' increased sensitivity to L2 intraword structures likely results from their greater need to attend to component letters and sequences during L1 decoding relative to Chinese speakers. This finding suggests that L1 experience can modulate learning of L2-specific regularities, with variable outcomes depending on how well the strategies acquired for the L1 can be applied to the L2 (see also Koda, 1990, 1993 for similar effects of L1 orthography and transfer on L2 reading comprehension strategies and Koda and Zehler, 2008, for review). The present study examines the possibility that effects of cross-linguistic similarity may be observed when languages overlap, not only in their orthographic systems as a whole, but in the orthographic forms of particular words.

Preliminary support for this possibility comes from another study by Koda (1989), who found that Japanese L2 learners whose L1 (Korean or Chinese) overlapped with one type of Japanese script (*kanji*, a logographic, meaning-based system), but not another (*hiragana*, a phonetic lettering system), outperformed learners with orthographically dissimilar L1s in learning vocabulary of both scripts. Furthermore, initial L1-similarity advantages for word learning compounded to yield later benefits for more complex tasks, such as reading comprehension. While these findings indicate that L1 similarity for a subset of L2 vocabulary can facilitate the acquisition of other L2-specific wordforms (potentially via transfer of phonological representations that map to both *kanji* and *hiragana*), learners with knowledge of logographic characters could have benefited from overlap in both orthographic form and meaning (akin to cognate facilitation). The present study therefore examines whether similar benefits of cross-linguistic influence on within-language transfer can be observed when a subset of novel words overlap with the L1 in sublexical properties alone.

Given that adult language learners' primary approach when they start learning a new language is typically to identify and reuse perceived similarities to their native language (Ringbom and Jarvis, 2011), it would be consequential to know how increased activation and use of L1 knowledge during early stages of learning affects learners' ability to later rely on regularities within the L2. One possibility is that identifying useful similarities between the L1 and L2 during initial acquisition will enhance the ability to learn and use similarities within the L2. For instance, learning L2 words that share orthographic features with the L1 could establish a stronger base of knowledge to be used as exemplars for subsequently learned words with L2-specific features. In addition to potential differences in the strength of L2 (exemplar) representations, the cognitive processes and strategies engaged while learning words that resemble the L1 could increase the salience and use of L2 regularities.

There are also reasons to expect that transfer between L2 words could instead be facilitated by the initial acquisition of wordforms that are *dissimilar* to the L1. For instance, the greater challenges associated with learning dissimilar words could serve as a form of desirable difficulty (Bjork and Bjork, 2011), which could elicit higher levels of involvement ( Craik and Lockhart, 1972; Laufer and Hulstijn, 2001; Rice and Tokowicz, 2020)

or motivation (Dörnyei and Ushioda, 2009; Dörnyei, 2019), resulting in deeper processing and greater sensitivity to L2-specific patterns. It may also be the case that words with less typical L1 orthography would elicit relatively less activation of L1 representations that could interfere with the identification and use of L2 features (e.g., Amengual, 2012, 2016; Goldrick et al., 2014). If so, we may observe greater within-L2 transfer after learning words with orthographic features that are uncommon in the L1.

In the present study, we introduce the concept of "bridge" words as a means to investigate cross-linguistic influences on transfer between novel vocabulary and the potential utility of bridge words for teaching learners about useful features of non-native words. Bridge words are defined as novel words that contain letter sequences that are common among the non-native vocabulary to facilitate subsequent learning. Acquiring a bridge word (e.g., *haner* in the current study) may make it easier to learn a similarly spelled "terminus" word (e.g., *hajer*) to which it is connected because of orthographic feature overlap. Some bridge words use letter sequences that are also common in the L1, which may make them easier to acquire, whereas other bridge words have orthographic forms that are uncommon in the L1. To examine the effect of L1 similarity on bridge words' utility, we designed contrasting sets of pseudowords and taught participants one of two word lists across two sessions. Participants were first taught bridge words comprised of letter sequences (i.e., bigrams) that were either typical (i.e., "Familiar;" e.g., *haner*, meaning "bride") or atypical of English words (i.e., "Unfamiliar;" e.g., *vobaf*, meaning "cloud"), followed by an immediate test where they produced the new word when cued with its meaning. Two weeks later, participants returned to learn terminus words that were related to their previously-learned bridge words (e.g., *hajer*, *tobaf*), and were again tested immediately. If we observe a general benefit for terminus word acquisition based on bridge word knowledge, it would suggest that learners are able to use orthographic similarities within the non-native vocabulary to facilitate subsequent learning. Critically, if we observe different effects of bridge words in the Familiar and Unfamiliar conditions, it would suggest that native language orthotactic typicality can modulate how knowledge specific to non-native words is used. Native language similarity may improve bridge-to-terminus transfer, by accentuating word-to-word similarity as a learning tool, or it may interfere, by hindering acquisition of non-native patterns.

## METHODS

### Participants

A power analysis to determine sample size was run with Monte Carlo simulations using the SIMR package in R for use with linear mixed effect models (Green and Macleod, 2016). An effect size for the influence of L1 orthographic bigram typicality on learning was obtained from word learning data in Bartolotti and Marian (2017), providing a fixed effect estimate of 10% on learning accuracy. Population mean and variance were obtained from pilot data. Power estimates were calculated for simulated

sample sizes from 20 to 40. Power  $>0.8$  was obtained with 30 participants, and power  $>0.9$  was obtained with 38 participants.

Sixty-five English-speaking adults initially participated after providing informed consent in accordance with the university's institutional review board, and were randomly assigned to learn Familiar or Unfamiliar word lists. Participants' language profiles were collected using the LEAP-Q (Marian et al., 2007). Non-verbal IQ was assessed using the matrix reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (PsychCorp, 1999). Verbal memory was assessed using the verbal paired associates test of the Wechsler Memory Scale III (Wechsler, 1997).

As the novel vocabulary used in the present study was controlled for orthographic wordform similarity to English (i.e., bigram and biphone probability), but not other languages, only native English-speakers with minimal second language knowledge were included in the final sample. Eligible participants had self-reported second language proficiencies (speaking, listening, and reading composite score) of less than 3 (corresponding to "low" proficiency) on a scale of 0–10 (ranging from "none" to "perfect"). This language knowledge criterion was applied prior to data analysis and excluded all bilinguals and multilinguals ( $N = 27$ ), yielding a final sample size of 38 participants (Familiar group  $N = 17$ , Unfamiliar group  $N = 21$ ). Participants in the Familiar and Unfamiliar groups did not differ in non-verbal IQ standard scores (Familiar  $M = 111.0$ ,  $SE = 0.55$ , Unfamiliar  $M = 110.5$ ,  $SE = 0.38$ ,  $t(28.1) = 0.17$ , n.s.) or verbal memory standard scores (Familiar  $M = 13.0$ ,  $SE = 0.15$ , Unfamiliar  $M = 13.4$ ,  $SE = 0.13$ ,  $t(32.6) = 0.45$ , n.s.). All participants were students at Northwestern University who completed the study in a classroom-like setting in exchange for extra credit.

## Materials

The Familiar and Unfamiliar word lists each contained 48 five-letter words with alternating consonants and vowels in CVCVC format (Q, Y, and X were not used in either language). Two versions of each word list were created, one per training session. Vocabulary items in the first list were used to examine L1 influences on learning, and were selected by evaluating 10,000 randomly generated non-words for English similarity. Though word lists were presented visually, psycholinguistic evidence suggests that phonological forms of words are co-activated even in response to unimodal orthographic inputs (e.g., Perfetti and Bell, 1991; Ferrand and Grainger, 1993; Grainger and Ferrand, 1994; Brysbaert et al., 1999; Van Wijnendaele and Brysbaert, 2002; Brysbaert and Van Wijnendaele, 2003; Grainger et al., 2006; Braun et al., 2009). English similarity was therefore determined based on both bigram and biphone probabilities. Phonological forms of each novel word were determined using the eSpeak speech synthesizer software, version 1.48.15 for Linux (Duddington, 2012). Pronunciations were obtained as IPA transcriptions using eSpeak's EN-US American English voice, and were translated from IPA to the CPSAMPA format (a modification of XSAMPA) for use with CLEARPOND (Marian et al., 2012). The orthographic and phonological forms of each novel word were used to obtain average bigram and biphone probabilities in English, and English similarity was

defined as a composite metric of  $z$ -transformed bigram and biphone probabilities.

To establish high and low English similarity thresholds, an English similarity percentile rank score was defined based on real English words. All five-letter English words in SUBTLEX-US (Brysbaert and New, 2009) with a frequency-per-million of 0.33 or greater were used to create the English similarity score. Each real word's score (i.e., average of  $z$ -transformed English bigram and biphone probabilities) was calculated and words were rank-ordered by English similarity. A High English similarity threshold was defined at the 20th percentile score, and 48 of the randomly generated novel words with scores above the threshold were selected for the first Familiar word list. A Low English similarity threshold was defined at the 99th percentile score, and 48 of the novel words with scores below the threshold were selected for the first Unfamiliar word list. Words in both lists were selected with the additional constraint of ensuring a balanced distribution of letters at word onset.

An additional 48 novel words in each condition (Familiar, Unfamiliar) were designed for use in the second session, which examined the effect of similarity to previously-learned words on new word learning. All new "terminus words" in the second session were substitution neighbors of a single item from that condition's "bridge word" list, learned in the first session. New terminus words were selected from a list comprising all non-word single-letter substitution neighbors of entries from the bridge word list (excluding duplicate entries, which were neighbors of multiple words in the bridge word list). In order to assess how well learners are able to utilize non-native patterns to learn other new words, English similarity was calculated for all generated entries and only new terminus words with scores below the Low English similarity threshold were selected for *both* the Familiar and Unfamiliar conditions. In other words, while the Familiar and Unfamiliar bridge word lists differed in English bigram/biphone probability for the first session, terminus words in the second session were equally dissimilar to English, thereby ensuring that effects of condition observed for terminus words could not be attributed to direct transfer from the L1. From this reduced list, 48 terminus words were randomly selected for each condition, with the constraints that each terminus word was a neighbor of a different word from the bridge word list and that the average English bigram/biphone probability did not differ between the second lists in each condition or between the bridge and terminus lists in the Unfamiliar condition (all  $ps > 0.1$ ; see **Supplementary Tables 1, 2** for bridge and terminus wordform statistics and stimuli).

All novel words were assigned a different English meaning for use during learning; the Familiar and Unfamiliar conditions both used the same list of 96 English words. To control for effects of individual novel-word—English-word pairings, two variants were created for each condition. The 96 English words were divided into A and B lists that each included equal numbers of concrete (e.g., "tree") and abstract (e.g., "idea") nouns (as determined by measures of imageability, see De Groot, 2006 for a similar approach). The two lists were matched for imageability, age of acquisition, and familiarity (Bristol norms) (Stadthagen-Gonzalez and Davis, 2006), as well as lexical frequency on

the SUBTLEX-US zipf scale (Brysbaert and New, 2009; Van Heuven et al., 2014) ( $p_s > 0.05$ ; see **Supplementary Tables 3, 4** for English words in lists A and B, as well as statistical comparisons of word characteristics in the two lists). For half of the participants in each language group, list A meanings were assigned to novel words in the bridge session and list B was used for the terminus session, while the other half of participants received list B meanings in the bridge session and list A meanings in the terminus session. In this way, each participant learned a translation of the same 96 English words (with each English word paired with a single novel word), with imageability, age of acquisition, familiarity, and lexical frequency controlled across the four list types (Familiar-Bridge, Familiar-Terminus, Unfamiliar-Bridge, Unfamiliar-Terminus). Lastly, in order to account for possible differences between groups in similarity between the novel wordforms and the wordforms of their direct English translations (e.g., a cognate effect or near-cognate effect), we confirmed that the number of novel word–English word pairs that had overlapping bigrams (e.g., a novel word “cohuz” paired with the English word “command”) did not differ between the Familiar and Unfamiliar Bridge word lists (2 and 0 out of 96, respectively) or between the Familiar and Unfamiliar Terminus word lists (1 out of 96 in both,  $p_s > 0.05$ ).

## Procedure

Participants learned the novel bridge and terminus word lists they were assigned over the course of two sessions spaced 2 weeks apart. In each session, each participant was given a sheet of paper containing all 48 novel bridge or terminus words and their meanings printed as paired associates (e.g., *haner*—bride). Participants were provided 16 min to silently learn as many words as they could, and were told that they would be tested immediately afterwards. While the use of a more structured task (e.g., timed presentation of individual word pairs) can be beneficial for isolating the mechanisms underlying effects on learning, the present study was designed to be an initial test of the hypothesis that similarity to native language words would modulate transfer of non-native knowledge. The use of carefully controlled word stimuli combined with a self-paced paired-associates task enabled us to assess the overall impact of native language similarity on non-native transfer without imposing constraints on learners' allocation of time to study individual words. This approach additionally allowed us to simultaneously test groups of participants in a classroom-like setting using a format commonly found in foreign language textbooks and study materials (see Prince, 1996; Laufer and Shmueli, 1997; Hermann, 2003; Webb, 2007 for similar approaches). The duration of the study phase was determined based on pilot data and prior studies utilizing similar paradigms (e.g., Pickering, 1982; Prince, 1996; Laufer and Shmueli, 1997; Webb, 2007). Following the study phase, participants were then given 6 min to write the matching novel word translations on a response sheet containing all 48 English meanings. The order of words was fixed across participants but randomized between learning and test. A research assistant later manually transcribed written responses onto a computer, which automatically scored participants' accuracy.

## Data Analysis

Response accuracy was calculated taking into account partially correct responses. Each correct letter in the correct position of a response scored 0.2 points, for a maximum score of 1. The effects of native and non-native word similarity on accuracy were analyzed with linear mixed effects-regression, using the lme4 package (Bates et al., 2014) in R (R Core Team, 2016). Models included fixed effects of Similarity Condition (Familiar, Unfamiliar) and Session (Bridge word, Terminus word), plus an interaction term. Imageability, familiarity, age of acquisition, and word frequency of the English translations were added as covariates. The models additionally included random intercepts for participants, word forms, and word meanings, allowing us to control for mean learning performance associated with individual participants and words. Models additionally included a by-participant random slope for Session and by-meaning random slopes for Session and Similarity (i.e., the “maximal” random effects structure<sup>1</sup>, Barr et al., 2013), allowing us to control for random variation in the fixed effects associated with individual participants and words. Significance of fixed effect estimates was evaluated using the Satterthwaite approximation for degrees of freedom. Follow-up comparisons on models' predicted marginal means (using Welch *t*-tests) also used the Satterthwaite approximation for degrees of freedom, and the Tukey correction for multiple comparisons.

## RESULTS

We found a significant interaction between Similarity and Session [*Estimate* = −13.35, *SE* = 4.93, 95% CI (−23.01, −3.69), *t*(51.65) = −2.71,  $p = 0.009$ ], as well as a main effect of Similarity [*Estimate* = 11.37, *SE* = 4.26, 95% CI (3.02, 19.72), *t*(43.98) = 2.67,  $p = 0.011$ ] and a marginally significant main effect of Session [*Estimate* = −4.44, *SE* = 2.54, 95% CI (−9.42, 0.54), *t*(58.32) = −1.75,  $p = 0.086$ ] (**Figure 1**). Follow-up comparisons on the model's predicted marginal means revealed that accuracy for the Familiar condition in the Bridge session  $M = 34.03$ , *SE* = 3.65, 95% CI (26.88, 41.19) was higher than for the Unfamiliar condition in the Bridge session  $M = 16.16$ , *SE* = 3.22, 95% CI (9.85, 22.48),  $z = -3.69$ ,  $p = 0.001$ , and higher than accuracy for either the Familiar condition  $M = 22.18$ , *SE* = 3.74, 95% CI (14.85, 29.52),  $z = 3.2$ ,  $p = 0.007$  or the Unfamiliar condition in the Terminus session  $M = 17.66$ , *SE* = 3.38, 95% CI (11.03, 24.29),  $z = 3.26$ ,  $p = 0.006$ . No other comparisons were significant.

The higher accuracy in the Bridge session for the Familiar condition compared to the Unfamiliar condition demonstrates a substantial benefit of native language similarity during self-directed vocabulary learning. However, the better learning observed for the Familiar condition did not carry through to the subsequent Terminus session, at which point there was no

<sup>1</sup>Note that no random slopes for Similarity were included for participants or word forms as each participant and word form was assigned to a single similarity condition (either Familiar or Unfamiliar). No random slope for Session was included for word form as each word form was presented in a single session (either bridge or terminus).

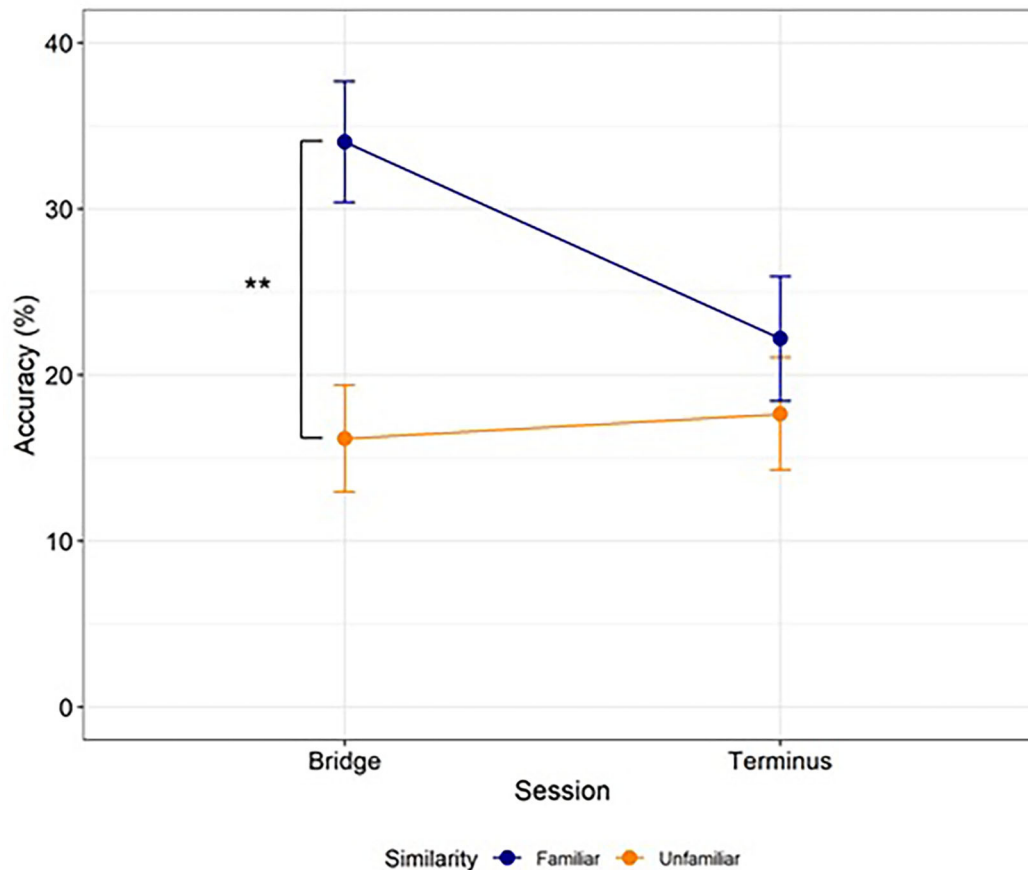


**FIGURE 1 |** Word learning accuracy. Learners in the Familiar condition (blue) acquired more words in the Bridge session than learners in the Unfamiliar condition (orange), providing evidence of a between-language similarity benefit. Accuracy decreased from the Bridge to the Terminus session for learners in the Familiar condition. Dots and error bars represent observed values and standard error, respectively, by participants. Lines represent the best fit linear mixed-effects regression model. \*\* $p < 0.01$ .

significant difference between word retrieval accuracy in the two groups.

The Terminus session contained entirely new vocabulary for participants to learn; all words were single letter substitution neighbors of words from the Bridge session (e.g., bridge word *haner* and terminus word *hajer*). To determine whether vocabulary that individuals learned in the Bridge session transferred to the Terminus session, we analyzed the data by first assigning each terminus word for each participant to one of three categories based on how well their substitution neighbors were learned during the Bridge session. Items in the Known Neighbor category were neighbors of bridge words that an individual got 4–5 out of 5 letters correct in the prior session. The Partly-Known Neighbor category included neighbors of bridge words with a score between 1 and 3 letters correct, and the Unknown Neighbor category included neighbors of bridge words that got a score of 0 letters correct. Note that items were assigned to Bridge-Knowledge conditions individually for each participant based on their performance in the Bridge session, and thus categories

have an unbalanced number of items [see **Table 1**;  $\chi^2(2) = 96.39$ ,  $p < 0.001$ ].

The model included fixed effects of Similarity Condition (Familiar vs. Unfamiliar) and Bridge-Knowledge (Known vs. Unknown Neighbor, and Known vs. Partly-Known Neighbor) plus interactions, as well as random intercepts for participant, word form, and word meaning, by-participant and by-form random slopes for Bridge-Knowledge, and by-meaning random slopes for Bridge-Knowledge and Similarity. Imageability, familiarity, age of acquisition, and word frequency of terminus words' English translations were entered as covariates.

We found a significant interaction between Similarity and Bridge-Knowledge [Known vs. Partly-Known contrast, *Estimate* = 15.02, *SE* = 6.29, 95% CI (2.69, 27.34),  $t(66.5) = 2.39$ ,  $p = 0.019$ ], but not the [Known vs. Unknown contrast; *Estimate* = 6.25, *SE* = 5.25, 95% CI (−4.04, 16.54),  $t(96.6) = 1.19$ ,  $p = 0.237$ ] and a main effect of Bridge-Knowledge [Known vs. Unknown contrast, *Estimate* = −11.69, *SE* = 3.12, 95% CI (−17.82, −5.57),  $t(96.8) = -3.74$ ,  $p < 0.001$ ; Known vs. Partly-Known contrast, *Estimate* = −11.35, *SE* = 3.45, 95% CI (−18.11, −4.6),  $t(78.9)$

**TABLE 1 |** Percentage of terminus words with known, partly-known, and unknown neighbors.

|                      | Known neighbors | Partly-known neighbors | Unknown neighbors |
|----------------------|-----------------|------------------------|-------------------|
| Familiar condition   | 29.2%           | 17.9%                  | 53.1%             |
| Unfamiliar condition | 11.0%           | 17.4%                  | 71.6%             |

Percentages reflect the average proportion of Terminus words with Known, Partly-Known, and Unknown Bridge word neighbors in the Familiar and Unfamiliar conditions. Each Terminus word's (e.g., *hajer*) category was determined on a participant-by-participant basis depending on how well its orthographically-related Bridge word (e.g., *haner*) was learned in the previous test session. A Terminus word was categorized as having a Known Neighbor if the participant correctly recalled 4–5 out of 5 letters of its corresponding Bridge word, as Partly-Known if they correctly recalled 1–3 letters, and as Unknown if they correctly recalled 0 letters. To determine the overall distribution of Terminus words in each category, the percentage of words with Known, Partly-Known, and Unknown neighbors was calculated for each participant and then averaged across participants within the Familiar and Unfamiliar conditions.

$= -3.30, p = 0.001$ ] (**Figure 2**). Follow-up comparisons on the model's predicted marginal means revealed that accuracy for Known Neighbor words in the Unfamiliar condition  $M = 30.75$  [ $SE = 5.37$ , 95% CI (20.0, 41.5)] was higher than for both Partly-Known Neighbors  $M = 12.68$  [ $SE = 4.35$ , 95% CI (3.8, 21.5)],  $t(45.4) = 3.49, p = 0.003$  or Unknown Neighbors  $M = 16.25$  [ $SE = 3.2$ , 95% CI (9.8, 22.7)],  $t(50.0) = 3.18, p = 0.007$ . In contrast, accuracy for Known Neighbor words in the Familiar condition  $M = 27.60$ , [ $SE = 4.65$ , 95% CI (18.2, 37.0)] did not differ from either Partly-Known Neighbors  $M = 24.55$ , [ $SE = 4.81$ , 95% CI (14.8, 34.3)],  $t(24.4) = 0.67, p = 0.781$  or Unknown Neighbors  $M = 19.36$ , [ $SE = 3.61$ , 95% CI (12.1, 26.7)],  $t(31.0) = 2.16, p = 0.094^2$ . Partly-Known Neighbors did not differ from Unknown Neighbors in either condition. These results show that learning a word in the Bridge session increased one's chances of learning its neighbor in the Terminus session, providing evidence that similarity to previously-learned novel words benefits later vocabulary acquisition. Critically, similarity to previously-learned words influenced the types of words that people learned in the Unfamiliar condition more than in the Familiar condition. The significant difference between Known Neighbors and both Partly-Known and Unknown Neighbors, but not between Partly-Known and Unknown Neighbors further suggests that complete acquisition of a bridge word was necessary for participants in the Unfamiliar condition to benefit from similarity to previously-learned words. Partially learning a bridge word did not result in any differences between the Familiar and Unfamiliar conditions.

## DISCUSSION

The goal of the present study was to determine how wordform similarity to the native language (as determined by bigram

<sup>2</sup>Similar results were obtained when Bridge word accuracy was instead entered as a continuous variable, showing a significant effect of Bridge-Knowledge in the Unfamiliar condition [ $Estimate = 10.65$ ,  $SE = 4.56$ ,  $t(49.8) = 2.34, p = 0.024$ ], but a marginal effect in the Familiar condition [ $Estimate = 7.71$ ,  $SE = 3.95$ ,  $t(83.8) = 1.95, p = 0.054$ ].

probability) influences acquisition of non-native vocabulary and sensitivity to non-native sublexical regularities. We found that although native language similarity provides short-term benefits, it can reduce reliance on non-native patterns during subsequent learning. Through continued use of an L2, the learner recognizes new patterns that determine how letters or sounds can combine to form words, and how words combine to form sentences. This process of extracting new patterns is also important for establishing continuous vocabulary learning, by ensuring that new words are accurately perceived and encoded in memory. Advanced L2 learners have been shown to benefit from L2 similarity during word learning (Stamer and Vitevitch, 2012; Ma et al., 2015), and in the current study, we found that similarity to other non-native words can also affect the earliest stages of vocabulary acquisition. Specifically, learning a word in the first session increased the likelihood that a similar word would be acquired in the subsequent session. Notably, while words that resembled the L1 were easier to learn at first, they had less of an influence on subsequent word learning. These results demonstrate the important roles of the native language, the burgeoning non-native vocabulary, and their interactions on new word learning.

Because of the way the new vocabulary in our study was designed, each word in the bridge session had a single substitution neighbor in the subsequent terminus session. These bridge-terminus word pairs allowed us to assess differences in word learning based on whether or not the learner already knew a similar word. Importantly, this is based not on intrinsic properties of the words, but instead on learners' idiosyncratic knowledge of patterns in the new word lists. Given the self-directed nature of the training session, the effect of similarity to previously-learned words that we observed may reflect how attention and study time were allocated to new words. Because overall accuracy did not improve between bridge and terminus sessions, the observed advantage for terminus words with already-acquired bridge neighbors comes at the expense of words with unlearned neighbors, consistent with prior self-directed word learning paradigms (Bardhan, 2010). The relative disadvantage for words with partly-learned neighbors may additionally result from the confusion that can occur between formally similar L2 words (e.g., the German words *Schafe*, meaning "sheep" and *schaffen*, meaning "create"; Laufer, 1988, 1989). Laufer (1988) conjectures that these types of "synform errors" may result from weak or unstable representations of L2 words in memory that could impair the learner's ability to distinguish between them or correctly map them to their corresponding meanings. Transfer from previously-learned wordforms may therefore have contrasting effects on subsequent learning depending on how well the initial words were learned, with facilitation from robustly encoded exemplars but interference from more unstable representations.

Notably, learners in the Familiar and Unfamiliar conditions differed in how much similarity to previously-learned words affected their continued learning. Even though bridge words in the first session were learned twice as well in the Familiar condition compared to the Unfamiliar condition, the effect of learning similar words in the terminus session was nearly twice as

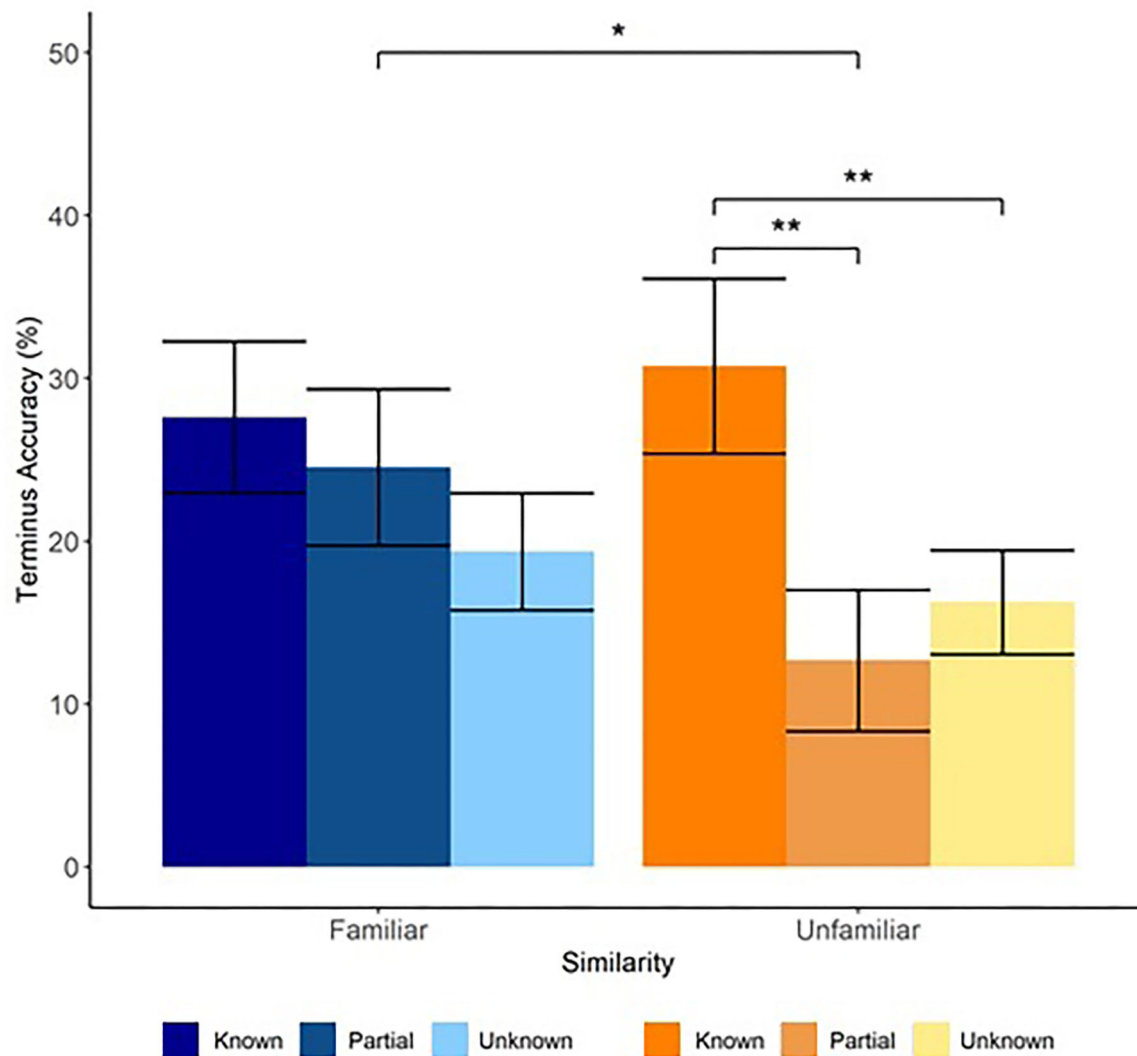


**FIGURE 2 |** Similarity to previously-learned novel words influences later acquisition. Word learning in the Bridge session affected the likelihood of learning its orthographic neighbor in the Terminus session. The effect of prior novel word learning was moderated by similarity to the native language. Accuracy in the Unfamiliar condition was higher for Known Neighbor words (dark orange) than for Partly-Known Neighbor words (orange) and Unknown Neighbor words (light orange); accuracy for Partly-Known and Unknown Neighbor words did not differ from each other. Accuracy in the Familiar condition did not differ between Known Neighbor words (dark blue), Partly-Known Neighbor words (blue), and Unknown Neighbor words (light blue). Error bars represent standard error (by participants). \* $p < 0.05$ ; \*\* $p < 0.01$ .

large for learners in the *Unfamiliar* condition. In the Unfamiliar condition, terminus words with learned bridge word neighbors were recalled with 2.65 times greater accuracy than words with unlearned bridge neighbors, compared to only a 1.65 times advantage in the Familiar condition. These terminus words in the second session were carefully designed to have equally low English similarity in both conditions, ensuring that this terminus word difference was due to effects of similarity to other non-native words, without confounding native and non-native word similarity. Together, these results indicate that although native language similarity provided an early benefit for word learning, it reduced the benefit of similarity to previously-learned non-native words in continued study.

Part of the task of learning a second language and achieving lexical competence involves building a foundation of L2

knowledge and a network of connections among L2 words and their meanings (Ellis and Beaton, 1993; De Groot, 2011), which can enhance the automaticity of L2 processing and minimize reliance on, and interference from, L1 knowledge (MacWhinney, 1997; Jiang, 2000). Connectionist models of bilingual language processing suggest that language selection and control can be accomplished over time via Hebbian learning and self-organizing representations that naturally cluster in language-specific ways due to greater feature overlap and co-activation of words within languages, than across languages (e.g., Shook and Marian's, 2013 *BLINCS* model). Such a system could allow bilinguals to rely on bottom-up inputs, such as orthographic or phonological features, to selectively activate the appropriate language based on learned regularities within each language. For instance, language-specific sublexical cues, such as letter and bigram frequencies,

can reduce the activation of cross-linguistic primes (Casaponsa and Duñabeitia, 2016; Dubey et al., 2018), and bilinguals can rely on language membership cues to guide lexical access (Grainger and Beauvillain, 1987; Vaid and Frenck-Mestre, 2002; Casaponsa et al., 2014) and speech production (Oganian et al., 2015). Participants in the current study who learned Familiar bridge words did not have orthographic cues that could reliably indicate language membership prior to lexical processing, which could have increased the activation of English representations relative to participants in the Unfamiliar condition. This may have stalled the process of linking new words into a coherent L2, interfering with transfer between the bridge and terminus words. In contrast, learners in the Unfamiliar condition were acquiring vocabulary that was unambiguously distinct from English. This distinction appears to be helpful in promoting extraction of non-native patterns to be used during new word learning.

The fact that the Familiar and Unfamiliar conditions did not differ in overall terminus word accuracy, however, may indicate that the two groups made use of different strategies or could have differed in other meaningful ways, such as in motivation, which has been shown to benefit word learning (Dörnyei and Ushioda, 2009; Dörnyei, 2019). For instance, the relative ease of learning bridge words that were similar to the L1 could have reduced motivation and effort in the Familiar condition, particularly during the second session when the task was unexpectedly more difficult. This could have elicited shallower processing of the terminus words during the word learning phase and consequently, reduced transfer from known bridge words. A complimentary interpretation would be that learners in the Unfamiliar condition benefited from the “desirable difficulties” (Bjork and Bjork, 2011) associated with learning more challenging bridge words. In language learning, retention in long term memory is generally improved when learning requires a greater depth of processing and involvement (Craik and Lockhart, 1972; Laufer and Hulstijn, 2001; Rice and Tokowicz, 2020), which can be instigated by material presented in a more difficult context (Schneider et al., 2002; Bjork and Kroll, 2015). Examples of desirable difficulties include repeated testing in place of passive study, or interleaving blocks of different word lists rather than blocked study (Schneider et al., 2002; Bjork and Kroll, 2015; Marecka et al., 2021). Our results suggest that difficulties caused by properties of the words themselves may also be targets for increasing long-term learning. Future research incorporating measures of motivation and/or manipulations of task engagement (e.g., through game-like formats, De Vos et al., 2019; Yang et al., 2020; see Derakhshan and Khatir, 2015 for review) could help elucidate the potential role of affective variables in determining the impact of cross-linguistic influence on transfer between non-native vocabulary.

In conclusion, we found that new vocabulary learning is affected by both similarity to one’s native language and similarity to other newly learned words. Whereas native language similarity has a beneficial effect early on, it may decrease sensitivity to non-native word patterns that support later learning. This suggests that cross-linguistic influence is modulated by interactions

between existing native and non-native word knowledge, and that initial similarity to the native language can have dynamically changing consequences over the course of novel word learning. This is because the words that one successfully learns early-on can influence the words that one acquires later, by driving attention toward new words that look more like already acquired ones. This suggests that cross-linguistic influences on initial vocabulary learning could potentially have cascading effects on the makeup of one’s later vocabulary. Overall, these results demonstrate the complex relationship between native and non-native vocabulary, where similarity can have variable consequences for learning.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northwestern University Institutional Review Board. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

Conceptualization and design: VM, JB, and AB. Data collection: JB and AB. Statistical analysis: JB and SH. Writing—original draft preparation: VM and JB. Writing—review and editing: SH, VM, and AB. Project administration and funding acquisition: VM. All authors contributed to the article and approved the submitted version.

## FUNDING

Research reported in this publication was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R01HD059858 to VM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## ACKNOWLEDGMENTS

The authors thank the members of the Northwestern University *Bilingualism and Psycholinguistics Research Group* for comments on this work.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.651506/full#supplementary-material>

## REFERENCES

- Amengual, M. (2012). Interlingual influence in bilingual speech: cognate status effect in a continuum of bilingualism. *Bilingualism* 15, 517–530. doi: 10.1017/S1366728911000460
- Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Front. Psychol.* 7:617. doi: 10.3389/fpsyg.2016.00617
- Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge: Harvard University Press.
- Atkinson, R. C., and Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *J. Exp. Psychol.* 1, 126–133. doi: 10.1037/0278-7393.1.2.126
- Bardhan, N. (2010). *Adults' self-directed learning of an artificial lexicon: the dynamics of neighborhood reorganization* (Ph.D. Thesis). University of Rochester, Rochester, NY, USA.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bartolotti, J., and Marian, V. (2014). Wordlikeness and novel word learning. *Proc. Annu. Meet. Cogn. Sci. Soc.* 36, 146–151. Available online at: <https://escholarship.org/uc/item/2j8652jh>
- Bartolotti, J., and Marian, V. (2017). Orthographic knowledge and lexical form influence vocabulary learning. *Appl. Psycholinguist.* 38, 427–456. doi: 10.1017/S0142176161000242
- Bartolotti, J., Marian, V., Schroeder, S. R., and Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Front. Psychol.* 2:324. doi: 10.3389/fpsyg.2011.00324
- Bates, D. M., Machler, M., Bolker, B. M., and Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:121811. doi: 10.18637/jss.v067.i01
- Bijeljac-Babic, R., Biardau, A., and Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Mem. Cogn.* 25, 447–457. doi: 10.3758/BF03201121
- Bjork, E. L., and Bjork, R. A. (2011). “Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning,” in *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society, 2nd Edn*, eds M. A. Gernsbacher, and J. Pomerantz (New York, NY: Worth), 59–68.
- Bjork, R. A., and Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *Am. J. Psychol.* 128, 241–252. doi: 10.5406/amerjpsyc.128.2.0241
- Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., and Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Hum. Brain Mapp.* 30, 1977–1989. doi: 10.1002/hbm.20643
- Brooks-Lewis, K. A. (2009). Adult learners' perceptions of the incorporation of their L1 in foreign language teaching and learning. *Appl. Linguist.* 30, 216–235. doi: 10.1093/applin/amn051
- Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Brysbaert, M., Van Dyck, G., and Van de Poel, M. (1999). Visual word recognition in bilinguals: evidence from masked phonological priming. *J. Exp. Psychol.* 25, 137–148. doi: 10.1037/0096-1523.25.1.137
- Brysbaert, M., and Van Wijnendaele, I. (2003). The importance of phonological coding in visual word recognition: further evidence from second-language processing. *Psychol. Belg.* 43, 249–258. doi: 10.5334/pb.1011
- Casaponsa, A., Carreiras, M., and Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: the impact of orthographic markedness. *Front. Psychol.* 5, 424. doi: 10.3389/fpsyg.2014.00424
- Casaponsa, A., and Duñabeitia, J. A. (2016). Lexical organization of language-ambiguous and language-specific words in bilinguals. *Q. J. Exp. Psychol.* 69, 589–604. doi: 10.1080/17470218.2015.1064977
- Casaponsa, A., Thierry, G., and Duñabeitia, J. A. (2020). The role of orthotactics in language switching: an ERP investigation using masked language priming. *Brain Sci.* 10:22. doi: 10.3390/brainsci10010022
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Dev. Psychol.* 32, 867–873. doi: 10.1037/0012-1649.32.5.867
- Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., and García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: the role of stimuli list composition. *J. Exp. Psychol.* 41, 614–635. doi: 10.1037/xlm0000065
- Comesaña, M., Perea, M., Piñeiro, A., and Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *J. Exp. Child Psychol.* 104, 22–33. doi: 10.1016/j.jecp.2008.10.004
- Comesaña, M., Soares, A. P., Sánchez-Casas, R., and Lima, C. (2012). Lexical and semantic representations in the acquisition of L2 cognate and non-cognate words: evidence from two learning methods in children. *Br. J. Psychol.* 103, 378–392. doi: 10.1111/j.2044-8295.2011.02080.x
- Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X
- Cunningham, T. H., and Graham, C. (2000). Increasing native English vocabulary recognition through Spanish immersion: cognate transfer from foreign to first language. *J. Educ. Psychol.* 92, 37–49. doi: 10.1037/0022-0663.92.1.37
- De Groot, A. M. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Lang. Learn.* 56, 463–506. doi: 10.1111/j.1467-9922.2006.00374.x
- De Groot, A. M., and Keijzer, R. (2000). What is hard to learn is easy to forget: the roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Lang. Learn.* 50, 1–56. doi: 10.1111/0023-8333.00110
- De Groot, A. M. B. (2011). *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York, NY: Psychology Press. doi: 10.4324/9780203841228
- De Vos, J. F., Schriefers, H., and Lemhöfer, K. (2019). Noticing vocabulary holes aids incidental language word learning: an experimental study. *Bilingualism* 22, 500–515. doi: 10.1017/S1366728918000019
- Derakhshan, A., and Khatir, E. D. (2015). The effects of using games on English vocabulary learning. *J. Appl. Linguist. Lang. Res.* 2, 39–47. doi: 10.36892/ijlls.v1i1.22
- Dixon, L. Q., Zhao, J., Shin, J. Y., Wu, S., Su, J. H., Burgess-Brigham, R., et al. (2012). What we know about second language acquisition: a synthesis from four perspectives. *Rev. Educ. Res.* 82, 5–60. doi: 10.3102/0034654311433587
- Dörnyei, Z. (2019). “Task motivation. What makes an L2 task engaging?,” in *Researching L2 Task Performance and Pedagogy: In Honour of Peter Skehan*, eds Z. E. Wen, and M. J. Ahmadian (Amsterdam: John Benjamins), 53–66.
- Dörnyei, Z., and Ushioda, E. (2009). *Motivation, Language Identity and the L2 Self*. Bristol: Multilingual Matters. doi: 10.21832/9781847691293
- Dressler, C., Carlo, M. S., Snow, C. E., August, D., and White, C. E. (2011). Spanish-speaking students' use of cognate knowledge to infer the meaning of English words. *Bilingualism* 14, 243–255. doi: 10.1017/S1366728910000519
- Dubey, N., Witzel, N., and Witzel, J. (2018). Script differences and masked translation priming: evidence from Hindi-English bilinguals. *Q. J. Exp. Psychol.* 71, 2421–2438. doi: 10.1177/1747021817743241
- Duddington, J. (2012). *eSpeak*. Available online at: <http://espeak.sourceforge.net> (accessed August 26, 2016).
- Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism* 18, 145–162. doi: 10.1017/S1366728913000722
- Eckman, F. R. (2004). From phonemic differences to constraint rankings: research on second language phonology. *Stud. Second Lang. Acquisit.* 26, 513–549. doi: 10.1017/S027226310404001X
- Ellis, N. C. (2008). *Usage-Based and Form-Focused Language Acquisition: The Associative Learning of Constructions, Learned Attention, and the Limited L2 Endstate*. London: Routledge/Taylor and Francis Group.
- Ellis, N. C., and Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Lang. Learn.* 43, 559–617. doi: 10.1111/j.1467-1770.1993.tb00627.x
- Ellis, N. C., and Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: putting language in good order. *Q. J. Exp. Psychol.* 49, 234–250. doi: 10.1080/713755604

- Ferrand, L., and Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bull. Psychonom. Soc.* 31, 119–122. doi: 10.3758/BF03334157
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *J. Phon.* 15, 47–65. doi: 10.1016/S0095-4470(19)30537-6
- Gathercole, S. E., and Masoura, E. V. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory* 13, 422–429. doi: 10.1080/09658210344000323
- Goldrick, M., Runnqvist, E., and Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychol. Sci.* 25, 1031–1036. doi: 10.1177/0956797613520014
- Grainger, J., and Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *Q. J. Exp. Psychol.* 39, 295–319. doi: 10.1080/14640748708401788
- Grainger, J., and Ferrand, L. (1994). Phonology and orthography in visual word recognition: effects of masked homophone primes. *J. Mem. Lang.* 33, 218–233. doi: 10.1006/jmla.1994.1011
- Grainger, J., Kiyonaga, K., and Holcomb, P. J. (2006). The time course of orthographic and phonological code activation. *Psychol. Sci.* 17, 1021–1026. doi: 10.1111/j.1467-9280.2006.01821.x
- Green, P., and Macleod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498. doi: 10.1111/2041-210X.12504
- Hakuta, K. (1982). Interaction between particles and word order in the comprehension and production of simple sentences in Japanese children. *Dev. Psychol.* 18, 62–76. doi: 10.1037/0012-1649.18.1.62
- Hamada, M., and Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Lang. Learn.* 58, 1–31. doi: 10.1111/j.1467-9922.2007.00433.x
- Hancin-Bhatt, B., and Nagy, W. (1994). Lexical transfer and second language morphological development. *Appl. Psycholinguist.* 15, 289–310. doi: 10.1017/S0142716400065905
- Helms-Park, R., and Perhan, Z. (2016). The role of explicit instruction in cross-script cognate recognition: The case of Ukrainian-speaking EAP learners. *J. English Acad.* 21, 17–33. doi: 10.1016/j.jeap.2015.08.005
- Hermann, F. (2003). Differential effects of reading and memorization of paired associates on vocabulary acquisition in adult learners of English as a second language. *TESL-EJ* 7, 1–16.
- Holcomb, P. J., Grainger, J., and O’rouke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *J. Cogn. Neurosci.* 14, 938–950. doi: 10.1162/089892902760191153
- Horst, M., Cobb, T., and Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Read. Foreign Lang.* 11, 207–223.
- Hulstijn, J. H. (1997). “Mnemonic methods in foreign language vocabulary learning: theoretical considerations and pedagogical implications,” in *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, eds J. Coady, and E. Huckin (Cambridge: Cambridge University Press), 203–224. doi: 10.1017/CBO9781139524643.015
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York, NY: Routledge. doi: 10.4324/9780203935927
- Jiang, N. (2000). Lexical representation and development in a second language. *Appl. Linguist.* 21, 47–77. doi: 10.1093/applin/21.1.47
- Karza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., and Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang.* 127, 46–54. doi: 10.1016/j.bandl.2012.11.007
- Kaushanskaya, M., and Marian, V. (2009). Bilingualism reduces native-language interference during novel-word learning. *J. Exp. Psychol.* 35, 829–835. doi: 10.1037/a0015275
- Kellerman, E. (1978). *Giving Learners a Break: Native Language Intuitions as a Source of Predictions About Transferability*. Working Papers on Bilingualism. Toronto, ON: Ontario Institute for Studies in Education, 59–92.
- Koda, K. (1989). The effects of transferred vocabulary knowledge on the development of L2 reading proficiency. *Foreign Lang. Ann.* 22, 529–540. doi: 10.1111/j.1944-9720.1989.tb02780.x
- Koda, K. (1990). The use of L1 reading strategies in L2 reading: effects of L1 orthographic structures on L2 phonological recoding strategies. *Stud. Second Lang. Acquisit.* 12, 393–410. doi: 10.1017/S0272263100009499
- Koda, K. (1993). Transferred L1 strategies and L2 syntactic structure in L2 sentence comprehension. *Modern Lang. J.* 77, 490–500. doi: 10.1111/j.1540-4781.1993.tb01997.x
- Koda, K. (1998). The role of phonemic awareness in second language reading. *Second Lang. Res.* 14, 194–215. doi: 10.1191/026765898676398460
- Koda, K., and Zehler, A. M. (2008). *Learning to Read Across Languages: Cross-Linguistic Relationships in First-and Second-Language Literacy Development*. New York, NY: Routledge. doi: 10.4324/9780203935668
- Laufer, B. (1988). The concept of ‘synforms’ (similar lexical forms) in vocabulary acquisition. *Lang. Educ.* 2, 113–132. doi: 10.1080/09500788809541228
- Laufer, B. (1989). A factor of difficulty in vocabulary learning: deceptive transparency. *AILA Rev.* 6, 10–20.
- Laufer, B., and Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Appl. Linguist.* 29, 694–716. doi: 10.1093/applin/amn018
- Laufer, B., and Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Appl. Linguist.* 22, 1–26. doi: 10.1093/applin/22.1.1
- Laufer, B., and Shmueli, K. (1997). Memorizing new words: does teaching have anything to do with it? *RELC J.* 28, 89–108. doi: 10.1177/003368829702800106
- Lin, A. M. Y. (2015). Conceptualising the potential role of L1 in CLIL. *Lang. Cult. Curriculum* 28, 74–89. doi: 10.1080/07908318.2014.1000926
- Liskin-Gasparro, J. E. (1982). *ETS oral Proficiency Testing Manual*. Princeton, NJ: Educational Testing Service.
- Lotto, L., and de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Lang. Learn.* 48, 31–69. doi: 10.1111/1467-9922.00032
- Ma, T., Chen, B., Lu, C., and Dunlap, S. (2015). Proficiency and sentence constraint effects on second language word learning. *Acta Psychol.* 159, 116–122. doi: 10.1016/j.actpsy.2015.05.014
- MacWhinney, B. (1997). “Second language acquisition and the competition model,” in *Tutorials in Bilingualism: Psycholinguistic Perspectives*, eds A. M. B. de Groot, and J. F. Kroll (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 113–142.
- MacWhinney, B., and Bates, E. (1989). *The Crosslinguistic Study of Sentence Processing*. Cambridge: Cambridge University Press.
- Majerus, S., Poncelet, M., Van der Linden, M., and Weekes, B. S. (2008). Lexical learning in bilingual adults: the relative importance of short-term memory for serial order and phonological knowledge. *Cognition* 107, 395–419. doi: 10.1016/j.cognition.2007.10.003
- Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Forýš-Nogala, M., Kutyłowska, K., et al. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition* 206:104477. doi: 10.1016/j.cognition.2020.104477
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS ONE* 7:e43230. doi: 10.1371/journal.pone.0043230
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- McLaughlin, J., Osterhout, L., and Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704. doi: 10.1038/nn1264
- Meade, G., Midgley, K. J., Dijkstra, T., and Holcomb, P. J. (2018). Cross-language neighborhood effects in learners indicative of an integrated lexicon. *J. Cogn. Neurosci.* 30, 70–85. doi: 10.1162/jocn\_a\_01184
- Meara, P. (1980). Vocabulary acquisition: a neglected aspect of language learning. *Lang. Teach. Linguist.* 13, 221–246. doi: 10.1017/S0261444800008879
- Melby-Lervåg, M., and Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: a meta-analysis of the correlational evidence. *J. Res. Read.* 34, 114–135. doi: 10.1111/j.1467-9817.2010.01477.x

- Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition* 108, 271–280. doi: 10.1016/j.cognition.2008.02.003
- Mori, Y. (1998). Effects of first language and phonological accessibility on kanji recognition. *Modern Lang. J.* 82, 69–82. doi: 10.1111/j.1540-4781.1998.tb02595.x
- Nation, I. P. (1982). Beginning to learn foreign vocabulary: a review of the research. *RELC J.* 13, 14–36. doi: 10.1177/003368828201300102
- Odlin, T. (1989). *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Oganian, Y., Conrad, M., Aryani, A., Heekeren, H. R., and Spalek, K. (2015). Interplay of bigram frequency and orthographic neighborhood statistics in language membership decision. *Bilingualism* 19, 578–596. doi: 10.1017/S1366728915000292
- Ortega, L. (2013). SLA for the 21st century: disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Lang. Learn.* 63, 1–24. doi: 10.1111/j.1467-9922.2012.00735.x
- Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C., and Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: a means for exploring the neurocognition of second language processing. *Lang. Learn.* 56, 199–230. doi: 10.1111/j.1467-9922.2006.00361.x
- Otwinowska, A. (2015). *Cognate Vocabulary in Language Acquisition and Use*. Bristol: Multilingual Matters. doi: 10.21832/9781783094394
- Otwinowska, A., Forys-Nogala, M., Kobosko, W., and Szewczyk, J. (2020). Learning orthographic cognates and non-cognates in the classroom: does awareness of cross-linguistic similarity matter? *Lang. Learn.* 70, 685–731. doi: 10.1111/lang.12390
- Otwinowska, A., and Szewczyk, J. M. (2019). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *Int. J. Bilingual Educ. Bilingualism* 22, 974–991. doi: 10.1080/13670050.2017.1325834
- Paivio, A., and Desrochers, A. (1981). Mnemonic techniques in second-language learning. *J. Educ. Psychol.* 73, 780–795. doi: 10.1037/0022-0663.73.6.780
- Papagno, C., Valentine, T., and Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *J. Mem. Lang.* 30, 331–347. doi: 10.1016/0749-596X(91)90040-Q
- Perfetti, C. A., and Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: evidence from backward masking and priming. *J. Mem. Lang.* 30, 473–485. doi: 10.1016/0749-596X(91)90017-E
- Pickering, M. (1982). Context-free and context-dependent vocabulary learning: an experiment. *System* 10, 79–83. doi: 10.1016/0346-251X(81)90070-1
- Prince, P. (1996). Second language vocabulary learning: the role of context versus translations as a function of proficiency. *Modern Lang. J.* 80, 478–493. doi: 10.1111/j.1540-4781.1996.tb05468.x
- PsychCorp (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.
- Puimège, E., and Peters, E. (2019). Learner's English vocabulary knowledge prior to formal instruction: the role of word-related and learner-related variables. *Lang. Learn.* 69, 943–977. doi: 10.1111/lang.12364
- Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Lang. Learn.* 53, 233–284. doi: 10.1111/1467-9922.00217
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.r-project.org> (accessed March 22, 2021).
- Rice, C. A., and Tokowicz, N. (2020). State of the scholarship: a review of laboratory studies of adult second language vocabulary training. *Stud. Second Lang. Acquisit.* 42, 439–470. doi: 10.1017/S0272263119000500
- Ringbom, H. (2007). Actual, perceived and assumed cross-linguistic similarities in foreign language learning. *AFinLAN vuosikirja*. doi: 10.21832/9781853599361
- Ringbom, H., and Jarvis, S. (2011). “The importance of cross-linguistic similarity in foreign language learning,” in *The Handbook of Language Teaching*, eds Long M. H., and Doughty C. J. (West Sussex, Wiley-Blackwell), 106–118. doi: 10.1002/9781444315783.ch7
- Roberts, J. A., Pollock, K. E., Krakow, R., Price, J., Fulmer, K. C., and Wang, P. P. (2005). Language development in preschool-age children adopted from China. *J. Speech Lang. Hear. Res.* 48:93. doi: 10.1044/1092-4388(2005/008)
- Rogers, J., Webb, S., and Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Lang. Teach. Res.* 19, 9–27. doi: 10.1177/1362168814541752
- Roodenrys, S., and Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *J. Exp. Psychol.* 28, 29–33. doi: 10.1037/0278-7393.28.1.29
- Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52. doi: 10.1016/S0010-0277(98)00075-4
- Schneider, V. I., Healy, A. F., and Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *J. Mem. Lang.* 46, 419–440. doi: 10.1006/jmla.2001.2813
- Service, E., and Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *J. Mem. Lang.* 32, 608–623. doi: 10.1006/jmla.1993.1031
- Shirai, Y. (1992). Conditions on transfer: a connectionist approach. *Appl. Linguist.* 3, 91–120.
- Shook, A., and Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism* 16, 304–324. doi: 10.1017/S1366728912000466
- Stadthagen-Gonzalez, H., and Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Beh. Res. Methods* 38, 598–605. doi: 10.3758/BF03193891
- Stamer, M. K., and Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism* 15, 490–502. doi: 10.1017/S1366728911000216
- Storkel, H. L., Armbrüster, J., and Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *J. Speech Lang. Hear. Res.* 49, 1175–1192. doi: 10.1044/1092-4388(2006/085)
- Storkel, H. L., and Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *J. Child. Lang.* 32, 827–853. doi: 10.1017/S0305000905007099
- Tekmen, E. A. F., and Daloglu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Lang. Ann.* 39, 220–243. doi: 10.1111/j.1944-9720.2006.tb02263.x
- Tonzar, C., Lotto, A., and Job, R. (2009). L2 vocabulary acquisition in children: effects of learning method and cognate status. *Lang. Learn.* 59, 623–646. doi: 10.1111/j.1467-9922.2009.00519.x
- Trévillat, M. C. (1996). Lexical learning and reading in L2 at the beginner level: the advantage of cognates. *Can. Modern Lang. Rev.* 53, 173–190. doi: 10.3138/cmlr.53.1.173
- Vaid, J., and Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French–English bilinguals. *Brain Lang.* 82, 47–53. doi: 10.1016/S0093-934X(02)00008-1
- Van Hell, J. G., and Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Lang. Learn.* 62, 148–171. doi: 10.1111/j.1467-9922.2012.00710.x
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* 67, 1176–1190. doi: 10.1080/17470218.2013.850521
- Van Wijnen, I., and Brysbaert, M. (2002). Visual word recognition in bilinguals: phonological priming from the second to the first language. *J. Exp. Psychol.* 28, 616–627. doi: 10.1037/0096-1523.28.3.616
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Lang. Learn.* 61, 219–258. doi: 10.1111/j.1467-9922.2010.00593.x
- Wang, T., and Saffran, J. R. (2014). Statistical learning of a tonal language: the influence of bilingualism and previous linguistic experience. *Front. Psychol.* 5:953. doi: 10.3389/fpsyg.2014.00953

- Webb, S. (2007). Learning word pairs and glossed sentences: the effects of a single context on vocabulary knowledge. *Lang. Teach. Res.* 11, 63–81. doi: 10.1177/1362168806072463
- Wechsler, D. (1997). *Wechsler Memory Scale III*. San Antonio, TX: Psychological Corporation.
- Wrembel, M. (2011). “Cross-linguistic influence in third language acquisition of voice onset time,” in *Proceedings of the 17th international congress of phonetic sciences*, eds W.-S. Lee, and E. Zee (Hong Kong: City University of Hong Kong), 2157–2160.
- Yang, Q. F., Chang, S. C., Hwang, G. J., and Zou, D. (2020). Balancing cognitive complexity and gaming level: effects of a cognitive complexity-based competition game on EFL students’ English vocabulary learning performance, anxiety and behaviors. *Comput. Educ.* 148:103808. doi: 10.1016/j.compedu.2020.103808
- Zahar, R., Cobb, T., and Spada, N. (2001). Acquiring vocabulary through reading: effects of frequency and contextual richness. *Can. Modern Lang. Rev.* 57, 541–572. doi: 10.3138/cmlr.57.4.541

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Marian, Bartolotti, van den Berg and Hayakawa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Equal Opportunity Interference: Both L1 and L2 Influence L3 Morpho-Syntactic Processing

Nawras Abbas<sup>1</sup>, Tamar Degani<sup>2</sup> and Anat Prior<sup>1,3\*</sup>

<sup>1</sup> Department of Learning Disabilities, University of Haifa, Haifa, Israel, <sup>2</sup> Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel, <sup>3</sup> Edmond J. Safra Brain Research Center for Learning Disabilities, University of Haifa, Haifa, Israel

We investigated cross-language influences from the first (L1) and second (L2) languages in third (L3) language processing, to examine how order of acquisition and proficiency modulate the degree of cross-language influences, and whether these cross-language influences manifest differently in online and offline measures of L3 processing. The study focused on morpho-syntactic processing of English as an L3 among Arabic-Hebrew-English university student trilinguals ( $n = 44$ ). Importantly, both L1 (Arabic) and L2 (Hebrew) of participants are typologically distant from L3 (English), which allows overcoming confounds of previous research. Performance of trilinguals was compared to that of native English monolingual controls ( $n = 37$ ). To investigate the source of cross-language influences, critical stimuli were ungrammatical sentences in English, which when translated could be grammatical in L1, in L2 or in both. Thus, the L3 morpho-syntactic structures included in the study were a mismatch with L1, a mismatch with L2, a Double mismatch, with both L1 and L2, or a no mismatch condition. Participants read the English sentences while their eye-movements were recorded (online measure), and they also performed grammaticality judgments following each sentence (offline measure). Across both measures, cross-language influences were assessed by comparing the performance of the trilinguals in each of the critical interference conditions to the no-interference condition, and by comparing their performance to that of the monolingual controls. L1 interference was evident in first pass sentence reading, and marginally in offline grammaticality judgment, and L2 interference was robust across second pass reading and grammaticality judgments. These results suggest that either L1 or the L2 can be the source of cross-language influences in L3 processing, but with different time-courses. The findings highlight the difference between online and offline measures of performance: processing language in real-time reflects mainly automatic activation of morpho-syntactic structures, whereas offline judgments might also involve strategic and meta-linguistic decision making. Together, the findings show that during L3 processing, trilinguals have access to all previously acquired linguistic knowledge, and that the multilingual language system is fully interactive.

**Keywords:** trilingualism, cross-language influence, morphosyntax, English as a foreign language, interference

## OPEN ACCESS

### Edited by:

Marianne Gullberg,  
Lund University, Sweden

### Reviewed by:

Greg Poarch,  
University of Groningen, Netherlands  
Anne L. Beatty-Martinez,  
McGill University, Canada

### \*Correspondence:

Anat Prior  
aprior@edu.haifa.ac.il

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 27 February 2021

**Accepted:** 16 April 2021

**Published:** 28 May 2021

### Citation:

Abbas N, Degani T and Prior A  
(2021) Equal Opportunity Interference:  
Both L1 and L2 Influence L3  
Morpho-Syntactic Processing.  
Front. Psychol. 12:673535.  
doi: 10.3389/fpsyg.2021.673535

## INTRODUCTION

Multilingualism can be considered a conventional feature of linguistic experience and maturity (Hammarberg, 2010). The growing prevalence of third language (L3) acquisition raises important theoretical considerations of how an additional language is represented and processed by multilingual speakers (Slabakova, 2017). While in second language (L2) acquisition learners rely solely on their experience with one language, in L3 acquisition two pre-existing systems of linguistic representations are available (Westergaard, 2019; Puig-Mayenco et al., 2020). Thus, investigating L3 acquisition allows researchers to clarify specific factors that might be confounded in L1 or L2 acquisition, such as how proficiency in a previous language might influence acquiring an additional language (Flynn et al., 2004).

Cross language influence is evident when acquisition or processing of one language is influenced by existing knowledge of other languages (Cenoz, 2001). Such influences can be facilitative, when structures of two languages are similar (positive transfer), but can also lead to language interference (negative transfer), in the presence of structural differences between the languages in question (Isurin, 2005; MacWhinney, 2005). There is a sizeable body of knowledge regarding how L1 can influence L2 processing (Hopp, 2010; Prior et al., 2017) and vice versa (Dussias and Sagarra, 2007; Degani et al., 2011), but our current understanding of how linguistic knowledge in L1 and/or L2 influences L3 learning and processing is far from being complete (Angelovska and Hahn, 2012; Rothman et al., 2019; Lago et al., 2020; Puig-Mayenco et al., 2020).

## Modulating Factors of Cross-Language Influences

One important dimension that has been emphasized as impacting cross-language influences in L3 learning and processing is the order of acquisition and/or proficiency in each of the background languages (Williams and Hammarberg, 2009). In addition, the typological similarity between each of the background languages and the L3 (Rothman and Halloran, 2013; Rothman, 2015) has also been identified as an important factor determining CLI in L3. Crucially, in much previous research these variables have been confounded or have been pitted against each other (e.g., Giancaspro et al., 2015; Puig-Mayenco et al., 2020). In the current study, we examine cross-language influence as a function of order of acquisition and/or proficiency in each of the background languages, irrespective of typological similarity, because for the examined population both L1 (Arabic) and L2 (Hebrew) are similarly typologically distant from the target L3 (English). Importantly, by using eye tracking as a measure of comprehension, we also examine the time-course of cross-language influences from each of the background languages, an issue which has received only very little attention in the extant literature.

When considering L3 processing, both L1 and L2 are potential sources of cross-language influences. However, there is ongoing debate regarding how these influences may play out, and whether one of the background languages becomes the “default supplier”

of cross-language influence (L1/L2) in L3 use. A strong preference for one of the previously acquired languages as providing cross-language influences for L3 has been suggested in some cases. For instance, some studies have identified L1 as the main source of cross-language influences in the acquisition of L3 in syntax and in lexicon (e.g., Gollan et al., 2002; Angelovska and Hahn, 2012). Hermas (2010) reported that among Arabic native speakers with L2 French and L3 English, the initial state of L3 syntax acquisition was influenced exclusively by the L1. Similarly, Lindqvist (2009) found that L1 was the main source of lexical influence on L3 French, among three groups with different background language combinations.

In contrast, many L3 acquisition studies have also identified cross-language influences that originate in the learner's L2 (e.g., Ringbom, 1987; Hammarberg, 2001; Bardel and Falk, 2007; Fallah et al., 2016). The “L2 Status Factor” theory explained that learners tend to activate the L2, rather than the L1, in L3 acquisition, because L2 is more similar to L3 with respect to the learning situation, age of onset, and degree of metalinguistic knowledge (Bardel and Falk, 2007; Falk and Bardel, 2011). Additionally, Bardel and Falk (2012), following neurolinguistic claims (Ullman, 2005), suggested that both L2 and L3 as non-native languages are stored in declarative memory, while the native language is stored in procedural memory. A study by Falk and Bardel (2011) in the domain of syntax, supported this hypothesis, by demonstrating that L2 superseded L1 as a source of both facilitation and interference in the L3. Specifically, using grammaticality judgment and a correction task, the study examined the placement of object pronouns in L3 German among two groups; L1 French-L2 English, and L1 English-L2 French. The results indicated that grammaticality judgments were influenced by participants' L2, and not L1, in both groups, suggesting that L2 had a stronger role than L1 in L3 acquisition (see also Angelovska and Hahn, 2012).

Lastly, recent models question the role of order of acquisition in granting privileged status to either L1 or L2 in cross-language influences on L3. For example, the *Scalpel Model* (Slabakova, 2017) argues against wholesale cross-language influence of previously acquired languages at the initial stages of acquisition, and instead posits that cross-language influences can come from the L1 or the L2 or both. Similarly, the *Linguistic Proximity Model* (Westergaard et al., 2017; Westergaard, 2019) suggests that in L3 acquisition, learners have access to all previously acquired languages, and that language acquisition is cumulative. In support of this claim, Westergaard et al. (2017) demonstrated that in a grammaticality judgment task in English (studied as a foreign language), monolingual Norwegian speaking children over accepted ungrammatical sentences, whereas Russian-Norwegian bilingual children and monolingual Russian speaking children noticed significantly more errors. However, the bilinguals scored lower than the L1 Russian speakers on grammatical sentences, suggesting the presence of interference from Norwegian. These results support the hypothesis that both previously acquired languages remain active and influence subsequent L3 acquisition, and that cross-language influences can be either facilitative or non-facilitative.

Several studies have found that either L1, L2, or both may contribute to cross-language influences in L3 acquisition (e.g., Flynn et al., 2004). For example, Bruhn de Garavito and Perpiñán (2014) found that speakers of L1 French – L2 English, at the initial stages of learning Spanish L3, rely in some situations on their L1 French grammar to interpret facts, and in other situations, they rely on their L2 English grammar. These findings suggest that L1 and L2 were both available and used whenever they facilitate processing of the input.

In addition to order of acquisition, individuals' proficiency in each of the background languages has also been cited as possibly contributing to the strength of L1 or L2 as sources of cross-language influences on L3. Specifically, high proficiency in a background language enables it to be influential in the acquisition of a new language (Williams and Hammarberg, 2009). For example, German was identified as the strongest source of cross-linguistic influence in acquisition of English, for monolingual German speaking adolescents but also for heritage speakers (of Turkish or Russian) who were immersed in German at the time of testing (Lorenz et al., 2019).

Finally, language typology has also been suggested as an important and influential factor in determining cross-language influences. The assumption is that the language that is more typologically similar to the L3, whether it is the L1 or the L2, will provide stronger influence during L3 acquisition and processing (Falk and Bardel, 2010; Angelovska and Hahn, 2012), as described by The *Typological Primacy Model* (Rothman, 2010, 2011, 2015).

The role of typology has been demonstrated in several studies. For example, Giancaspro et al. (2015) found that speakers of English and Spanish were dominantly influence by Spanish when learning L3 Brazilian Portuguese, regardless of whether Spanish was their L1 or L2. Analogous findings, of stronger influence from the typologically closer language, have also been demonstrated in the lexical domain (e.g., Ringbom, 1987; Poarch and Van Hell, 2014).

Importantly, studies that investigate the interplay of these various factors are often limited by confounds among them (Ecke, 2015; Rothman, 2015). Thus, many studies on L3 processing investigated the use of an L3 after the acquisition of an L2 which is more similar to L3 than is the L1 (e.g., De Angelis and Selinker, 2001). In such studies the effects of order of acquisition cannot be separated from those of typological similarity (Ecke, 2015). A few studies have tried to disentangle such combined effects, but provided mixed results. Cenoz (2001), for example, investigated bilingual speakers of Spanish and Basque learning English as an L3. Spanish is typologically more similar to English than Basque, but the results showed greater cross-language influences from Spanish when learning L3 English only when Spanish was the learner's L2, not when it was the L1. This finding demonstrated that beyond language typology, the L2 has an additive effect on cross-linguistic influence (see also Bardel and Falk, 2007, for similar results). On the other hand, two studies by Singleton and O'Laoire (2004, 2005) demonstrated that in the lexical domain the typology factor was stronger than the L2 status factor. Specifically, English, which is typologically closer to French in lexical terms, was the dominant source of cross-language influences in learning French as an L3, both for

English-L1–Irish-L2 bilinguals, and for bilinguals with both Irish and English as their L1s.

The results described above emphasize the difficulty of investigating cross-language influence in L3 processing, and the unique challenge of separating the impact of various factors. Thus, studies that have directly contrasted typology and order of acquisition as determining factors for cross-language influence have not reached a coherent conclusion – with some results identifying typology as the critical factor, and others identifying order of acquisition. The current study was designed to further investigate this issue, in a design that effectively neutralizes the typological factor, by studying L3 processing in trilinguals for whom both L1 and L2 are typologically *distant* from the L3.

## The Current Study

The main goal of this study is to examine whether L1 or L2 can be identified as an exclusive source of cross-language influence, or whether the entire linguistic repertoire is activated when processing L3 morpho-syntax. The specific population and methods we adopted allow us to complement previous research in several important ways.

Participants in this study are Arabic-Hebrew-English (AHE) university student trilinguals in Israel. This population expands upon previously studied samples in three aspects – the specific language combination, the ubiquity of trilingualism, and the level of proficiency. In most research conducted on L3 processing, the linguistic background included L1 or L2 (or even both) that are typologically similar to L3, and all three languages often belonged to the same language family (often Indo-European, Ecke, 2015). In contrast, the current study focuses on trilinguals whose background languages (L1-Arabic and L2-Hebrew) are Semitic, and whose L3 (English) is Indo-European. Moreover, each of the three languages is written in a different script, such that when reading English there is no orthographic overlap with either the L1 or the L2. Thus, in the current study, the language typology factor is neutralized since both L1 and L2 are typologically distant from L3.

Participants in the current study are also recruited from a large population of trilingual speakers. Many previous studies of trilingualism have focused on individuals who have self-selected to become multilinguals by studying additional languages (e.g., Lindqvist, 2009; Ecke and Hall, 2013). However, the current study extends the literature to test individuals who have become trilingual due to their social-educational context. All native Arabic speakers in Israel study both Hebrew (which is also the majority societal language) and English (as a foreign language) from early elementary school (age 8–9). Research with these learners is important, because it allows us to test the generalizability of previous findings in wider populations. Recently, several studies have examined non-self-selected individuals, by comparing monolingual and bilingual learners acquiring an additional language (Fallah et al., 2016; Westergaard et al., 2017; Hopp, 2019; Lorenz et al., 2019), but these all tested children who were at relatively early stages of L3 acquisition.

Given the socio-educational system in Israel described above, native Arabic speaking university students are moderately

proficient in both the L2 and the L3 (Prior et al., 2017). Specifically, at the time of testing, participants are partially immersed in the L2, in which they are conducting their studies, and are using L3 on a daily basis (see participant description below). Much previous theoretical interest has focused on early L3 acquisition, to identify the source of transfer in the initial state of learning (Rothman et al., 2019). Accordingly, in a recent systematic review of L3 learning, Puig-Mayenco et al. (2020) identified 40% of studies focusing on beginners, and the remainder as testing "post-beginners," but they acknowledge that this is a very wide category. An examination of the studies included in the review shows that a much lower percent actually tested individuals who had been using the L3 for an extended period of time (over 10 years in the current study). Here, however, we chose to investigate intermediate proficiency trilinguals, who habitually use all three languages, to reach a better understanding of how cross-language influences continue to impact L3 processing beyond the initial stages of acquisition.

The current study also differs from previous research in our approach to selecting language materials. Most previous studies identified one or two syntactic structures, that either differed in the overlap with the L1 and L2 of a single group of participants (e.g., Hopp, 2019), or they included two groups with different L1/L2 constellations (termed Mirror-Image groups by Puig-Mayenco et al., 2020), and focused on a single structure (e.g., Falk and Bardel, 2011; Cabrelli-Amaro et al., 2015). In the current study, we adopt a different approach. The study includes a single group of trilingual participants, Arabic-Hebrew-English trilinguals, who are compared with a control group of monolingual native English speakers. Thus, the target language is English for all participants. For trilinguals, English is the L3, L1 is always Arabic and L2 is always Hebrew. We further focus exclusively on interference in morpho-syntactic processing, or "non-facilitative" transfer. Specifically, critical items are always ungrammatical in English the L3, but could be grammatical in L1, L2 or both (for a somewhat similar approach see Westergaard et al., 2017). Accordingly, we define 4 conditions of syntactic overlap: structures in L3 that mismatch both L1 and L2 (which share a similar structure), structures that mismatch either L1 or L2 (but are shared across English and the other language), and structures which are common across all 3 languages (deemed control). For each such condition, we identified at least 3 syntactic structures in English (and after pre-testing, at least 2 remained for full analysis). Note that this method by definition includes different syntactic structures in the 4 experimental conditions, and these may differ in their basic ease or difficulty of acquisition/processing in English. To control for these potential baseline differences, our study therefore includes a control group of monolingual native English speakers, whose performance across the experiment serves as the baseline to which trilingual performance is compared. Finally, the critical stimuli are always presented as ungrammatical sentences in English. Cross-language influence is probed due to the fact that the ungrammatical structure presented in English would be grammatical in participants' L1, L2 or both. Our reasoning is that if there is indeed interfering cross-language influence from these languages, participants will

find it more difficult to identify the English critical sentences as ungrammatical.

Finally, the current study includes both online and offline measures of morpho-syntactic processing, by utilizing both recording of eye-movements during reading, and post-sentence grammaticality judgments. When overt decision tasks are used to study cross-language influences in L3 processing (e.g., Sanz et al., 2015; Slabakova and Garcia Mayo, 2015; Westergaard et al., 2017), participants normally wait to achieve a fairly high threshold level of certainty prior to responding. In contrast, the eye-movement record provides a window into the moment-by-moment processes underlying language comprehension (Dussias, 2010; Marinis, 2010; Sedivy, 2010). Recording eye-movements during reading provides a millisecond-precise report of the readers' immediate syntactic processing. It also provides an extremely rich data set, and may be used to determine when (e.g., during the first or second pass through a sentence) and where exactly in a sentence processing difficulty occurs, as well as how the reader deals with such difficulty (e.g., by rereading / fixating for longer durations / regressive saccades to an earlier point in the sentence) (Conklin and Pellicer-Sánchez, 2016).

In the current study, we combined recording of eye-movements during reading L3 sentences, with a post-sentence judgment of whether it was grammatically well formed in English. This allows us to investigate ongoing interference during processing, as well as more meta-linguistic processes of offline judgments. However, it is important to note that some previous research has demonstrated that incorporating grammaticality judgments invokes greater strategic processing as well as greater sensitivity to reading patterns during online reading (Godfroid and Winke, 2015; Keating and Jegerski, 2015). Thus, we acknowledge that some of the reading patterns identified in the current study might not be perfectly aligned with those evident during naturalistic reading, when readers are not simultaneously engaged in an additional task.

The current study aims to examine whether cross-language influences in L3 morpho-syntactic processing can be identified from both the L1 and the L2 when typological similarity is neutralized. We hypothesize that both L1 and L2 are potential sources for interference in L3 processing, as suggested by the theoretical stance of the Linguistic Proximity Model and the Scalpel Model. Accordingly, we predict significant interference from L1 and from L2 when there is a mismatch in syntactic structure with the L3. Further, we hypothesize that interference might be increased when L3 differs from both background languages, suggesting that the degree of structural mismatch can modulate cross-language influences.

A second aim of the current study is to test whether proficiency and/or order of acquisition modulate cross-language influences. In particular, we ask whether cross-language influence from the more dominant language L1 is expressed earlier in the time course of processing than is cross-language influence from the less proficient L2. Early and late eye movement measures may be revealing in this respect. Finally, the combination of online and offline measures employed in the current study will allow us to test whether the impact of cross-language influences on real-time

processing difficulty is similar to that expressed in metalinguistic based judgments.

## METHOD

### Participants

Fifty-three Arabic-Hebrew-English trilinguals (39 females, mean age 20.6) who were first year bachelor's degree students at the University of Haifa participated in the study. Previous research shows that this population is most proficient in L1, then in L2, and least proficient in L3 (Prior et al., 2017). This dominance profile was verified using objective and subjective proficiency measures in each language (see details below, and **Table 1** for participant characteristics). Participants grew up in exclusively Arabic speaking homes and schools. They started formal instruction in Hebrew at age 8 (2nd grade), had some exposure to Hebrew as the majority language in Israel, and at the time of data collection were immersed in college classes in Hebrew. Participants started formal instruction in English at age 9 (3rd grade), and had limited exposure to the language through media (music, television, film). Participants had no history of neurological or psychiatric deficits, learning or language disability and had intact or corrected vision. Nine participants were later excluded for not matching the required criteria, such that the final set of trilinguals included 44 participants (36 females, mean age  $M = 20.59$ ,  $SD = 1.46$ , range 19–27, Parental education  $M = 14.6$  years,  $SD = 4.1$ ). Participants were recruited through advertisements and received course-credit or payment for participation.

In addition, 37 monolingual native English speakers, students at the University of Wisconsin in Madison, participated in this study (33 females, mean age  $M = 20.02$ ,  $SD = 1.46$ , range 18–25). They were recruited as a control group for the experimental task in order to set the baseline performance accuracy and reading times across the interference conditions (see detailed description below). All participants gave informed consent to take part in the study. The study was approved by the University of Haifa Ethics Committee.

**TABLE 1 |** Trilingual participant characteristic ( $N = 44$ ).

|                                      | L1 (Arabic)               | L2 (Hebrew)               | L3 (English)              |
|--------------------------------------|---------------------------|---------------------------|---------------------------|
| Age began study                      | N/A                       | 8.27 (1.67)               | 8.50 (1.22)               |
| Current exposure (0–10 scale)*       | 7.17 (2.08) <sup>a</sup>  | 5.18 (1.84) <sup>b</sup>  | 3.55 (2.12) <sup>c</sup>  |
| Self-rated Proficiency (0–10 scale)* | 9.73 (0.49) <sup>a</sup>  | 8.02 (1.23) <sup>b</sup>  | 6.30 (1.53) <sup>c</sup>  |
| Semantic Fluency Task*               | 23.20 (5.94) <sup>a</sup> | 13.36 (6.92) <sup>b</sup> | 11.86 (5.08) <sup>b</sup> |
| Phonemic Fluency Task <sup>1**</sup> | 17.16 (5.15) <sup>a</sup> | 13.50 (5.66) <sup>b</sup> | 15.45 (4.35) <sup>c</sup> |

\* $P < 0.001$ ; Means in the same row with different superscript letters differ from each other significantly. Specifically, there were no significant differences in age of acquisition and in semantic fluency between L2 and L3.

\*\* $P < 0.05$ ; Ratings of proficiency and exposure were averaged across productive and receptive oral and written language use.

<sup>1</sup> Different phonemes were used across languages, informed by previous research. However, norming data collected in our lab after data collection of the current study revealed that these were not well matched across languages, with the English phonemes generating more responses than the Hebrew ones, which explains why these scores do not align well with the participants' expected language profile.

## Materials

**L3 (English) Sentence Processing Task.** Critical sentences were ungrammatical sentences in English that included a violation in one of four types of constructions: (1) Similar in Arabic and Hebrew, but different in English (Double mismatch, causing interference from both L1 and L2); (2) Similar in Hebrew and English, but different in Arabic (L1 mismatch, causing L1 interference); (3) Similar in Arabic and English, but different in Hebrew (L2 mismatch, causing L2 interference); and (4) Similar in Arabic, Hebrew, and English (control, no interference). Examples for these conditions are presented in **Table 2** (see **Supplementary Materials 1** for full materials). For each condition, we identified 3 potential structures (see **Supplementary Materials 2**, for further examples and explanations), and constructed 5 sentences for each structure for a total of 15 sentences per condition<sup>1</sup>. Critical sentences were constructed by considering that reliance on syntactic rules of Arabic, Hebrew, or both, may lead to an error in judging the grammaticality of English sentences. If translated word by word, the ungrammatical sentences presented in English would be grammatical in one or both of the background languages (L1 and L2), depending on the condition. For instance, the English sentence "I am planning to buy dog\* for my son's tenth birthday" is ungrammatical in English because the indefinite article is omitted. However, the participants might find it difficult to detect this violation, because if directly translated into either Arabic or Hebrew, it would be grammatical, since neither language has indefinite articles. To deal with the diglossic nature of Arabic (Saiegh-Haddad and Henkin-Roitfarb, 2014), we only selected structures that are shared between the spoken and the written variants of Arabic.

These 60 critical (ungrammatical) sentences were complemented by 60 (grammatical) filler sentences, constructed with no special constraint on cross-language influences. Both grammatical and ungrammatical sentences were simple active sentences, including high frequency vocabulary appropriate to participants' proficiency level in English, as ascertained by a pre-test in which 31 Arabic-Hebrew-English trilinguals rated sentence grammaticality, and verified general familiarity with the vocabulary included in the sentences. Sentence length ranged from 10 to 14 words, and was matched across conditions [ $F(3,55) = 2.11$ ,  $p = 0.109$ ]. Critical words (the words at which the grammatical violation is evident) were preceded and followed by at least two content words.

To pre-test materials, 27 monolingual native English speakers, students at the University of Wisconsin in Madison rated the grammaticality of the sentences and identified the error in the ungrammatical sentences. These participants did not take part in the main experiment. The rating task was performed online, with each student rating 60 sentences, half of which were grammatical. These ratings, in concert with the performance of the native English speakers on the experimental task, were used to determine the baseline difficulty of the structures and to select the final set of structures, such that they were matched for difficulty across conditions for native English speakers (see "Results" section below).

**TABLE 2 |** Examples of experimental materials: selected syntactic structures from the different mismatch conditions in Arabic, Hebrew, and English.

| Condition                                  | Construction                                      | L3 (English)                                                                                                  | L1 (Arabic)                                                                         | L2 (Hebrew)                                                                                       |
|--------------------------------------------|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| L1 mismatch / L2 similar                   | Possessive Marking                                | My classmate Adam always copies <u>*Sara homework</u> and the teacher never knows. [ <u>Sara's homework</u> ] | يقوم زميلي آدم دائماً بنسخ واجبات <u>سارة</u> المدرسية والمعلمة لا تعرف أبداً.      | החבר שלי אדם תמיד מעתיק את שיעורי הבית <u>שרה</u> והמורה אף פעם לא יודעת.<br>(שיעורי הבית של שרה) |
|                                            | Definite article omission in the superlative form | The coach likes Fadi because he is <u>*fastest</u> player in our team. [ <u>the fastest</u> ]                 | المدرب يحب فادي لأنه <u>أسرع لاعب في فريقنا</u> .                                   | המאמן אוהב את פאדי כי הוא <u>השחקן מהיר ביותר*</u> בקבוצה שלנו.<br>(המהיר ביותר)                  |
| L1 similar / L2 mismatch                   | Comparative Form                                  | My sister's hair is <u>*more long</u> than my hair which is really short. [ <u>longer</u> ]                   | شعر أختي <u>*طويل أكثر</u> من شعري القصير جداً.<br>(أطول)                           | השיער של אחותי <u>ארוך יותר</u> מהשיער שלי שהוא ממש קצר.                                          |
|                                            | Superlative Form                                  | Everyone knows that I'm the <u>*most rich</u> in this neighborhood. [ <u>richest</u> ]                        | يعلم الجميع أنني <u>*الأكثر غنى</u> في هذا الحي.<br>(الأغنى)                        | כולם יודעים שאני <u>העשיר ביותר</u> בשכונה הזאת.                                                  |
| Double Mismatch: L1 mismatch / L2 mismatch | First person prodrop                              | Selena won't talk to me even though <u>*visited</u> her last night. [ <u>I visited</u> ]                      | سلينا لن تتكلم معي على الرغم من أنني زرتها في الليلة الماضية.                       | סלינה לא מוכנה לדבר איתי למרות ש <b>ביקרתי</b> אותה אתמול בלילה.                                  |
|                                            | Copula Omission                                   | Ahmad won't come with us because he <u>*sick</u> and tired today. [ <u>is sick</u> ]                          | أحمد لن يأتي معنا لأنه <u>مريض</u> ومتعب اليوم.                                     | אחמד לא יבוא איתנו כי הוא <u>חולה</u> ועייף היום.                                                 |
|                                            | Indefinite article omission                       | I am planning to buy <u>*dog</u> for my son's tenth birthday. [ <u>a dog</u> ]                                | أخطط لشراء <u>كلب</u> في عيد ميلاد ابني العاشر.                                     | אני מתכננת לקנות <u>כלב</u> ליום ההולדת העשירי של הבן שלי.                                        |
| Control: No mismatch                       | Verb-time expression agreement                    | Yesterday, the students in my class <u>*will go</u> to Miami's best beach. [ <u>went</u> ]                    | بالأمس، <u>*سيذهب</u> الطلاب في صفي إلى أفضل شاطئ في ميامي.<br>(ذهب)                | אתמול התלמידים בכיתה שלי <u>*ילכו</u> לחוף הכי טוב במיאמי.<br>(הלכו)                              |
|                                            | Quantifier-noun agreement                         | Last week at the park, three <u>*dog</u> followed me, and I got scared. [ <u>dogs</u> ]                       | في الأسبوع الماضي في الحديقة شعرت بالخوف عندما تبعتني ثلاثة <u>*كلب</u> .<br>(كلاب) | בשבוע שעבר בפארק נבהלתי כאשר שלושה <u>*כלב</u> עקבו אחריי.<br>(כלבים)                             |

**Language Proficiency Assessment.** Participants' language profile was verified using both an objective verbal fluency task, and self-report measures derived from a detailed language history questionnaire, as detailed below.

**Verbal fluency tasks** (Gollan et al., 2002; Kavé, 2005). Participants were asked to produce in one minute as many words as possible within a given language for each of two semantic categories and each of two phonemes. In the semantic fluency task, three pairs of semantic categories (including one wide and one narrow category) were used and rotated randomly across the three languages for each participant: Animals and sports, fruits and occupations, and clothes and furniture (Gollan et al., 2002; Kavé, 2005). In the phonemic fluency task, a different pair of phonemes was used in each language in the following order: [c] and [r] were used in Arabic, [b] and [f] were used in Hebrew, [α] and [f] were used in English.

The order of languages was held constant across participants, so that both tasks (phonemic and semantic fluency tasks) were first administered in Arabic, then in Hebrew, and finally in English. The order of the tasks was counterbalanced, such that

participants were randomly assigned to either complete the phonemic fluency task first and the semantic fluency task second, or vice versa. However, the same order was administered in the three languages for each participant (e.g., if in Arabic the semantic task was administered first, and the phonemic task was second, the same order was retained same in Hebrew and in English).

**Language History Questionnaire.** Participants completed an Arabic translation of the Language Experience and Proficiency Questionnaire (a modified version of the LEAP-Q, Marian et al., 2007) in which they provided self-ratings of language use, language exposure, and language proficiency (across speaking, understanding spoken language, and reading) in all acquired languages.

## Procedure

Arabic-Hebrew-English trilinguals performed the experimental task, in addition to language proficiency tests at the University of Haifa in a single session. The tasks were administered in the following order: English Sentence Processing task (including

eye movement recording and post sentence grammaticality judgment task), Semantic and Phonemic fluency tasks, and then LEAP-Q (Marian et al., 2007). The order of test administration was held constant, except that the order of the fluency tasks was randomized. The entire experimental session lasted approximately an hour and a half.

Native English monolinguals completed the experimental task at the University of Wisconsin in Madison. They completed the identical English Sentence Processing task in a similar setting in Madison, except that the proper names included in the stimuli were English names and not Arabic (e.g., while the names “Ahmad” and “Yasmine” were used in the stimuli of the experimental group, the names “David” and “Jasmine” were used for the English speakers). They filled out a screening form to verify that inclusion criteria were met (monolingual speakers with no history of neurological or psychiatric disorder, learning or language disability and intact or corrected vision).

**English Sentence Processing Task.** Eye movements were recorded using an Eye Link 1000 eye tracker, which was tower-mounted in Haifa and desktop-mounted in Madison. Data were recorded monocularly from the pupil of the right eye at a sampling rate of 1,000 Hz. Chin and forehead rests were used to minimize head movement. Prior to the beginning of the experimental task, the eye-tracker was calibrated for each participant using a nine-point calibration grid, followed by a validation check. Then, the participants were presented with written instructions on the screen in their native language (Arabic or English). The instructions were followed by a practice block of 4 trials, and then by two experimental blocks of 60 trials each. The order of the sentences was set in the practice block and randomized in the experimental blocks.

Screen resolution was set at  $1024 \times 768$  pixels, and sentences were presented in black Courier New 14-point font on a white background. Sentences were left justified, and before each sentence, a one-point calibration check on the left side of the screen was conducted to ensure that participants consistently began reading the sentences at the leftmost point. Trials were terminated when participants fixated a gaze-contingent box at the right bottom corner of the screen when they finished reading each sentence. Following each sentence, a question mark appeared in the middle of the screen and participants provided their grammaticality judgment by button press (right key for a grammatically correct sentence, left key for a grammatically incorrect sentence). Participants were instructed to use only grammaticality as the basis for their judgment, and were also

instructed to respond as quickly and as accurately as possible. Feedback (smiley face/sad face on the screen) was provided in the practice block, but no feedback was given in the experimental blocks.

Participants were given a short break between the experimental blocks, and could also take a break at any point in the task between trials when necessary. The eye-tracking task took about 50 min to complete.

## RESULTS

### Equating Baseline Performance – Subset Selection

As mentioned earlier, data from monolingual native English speakers was collected in order to gauge the processing difficulty of the various target structures, and to characterize the baseline complexity of processing each structure in the absence of any cross-language influence. Preliminary examination of the accuracy in the grammaticality judgment of the native English control group revealed, however, differences across experimental conditions [ $F(3, 108) = 27.18$ ,  $MSE = 0.46$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.43$ ]. To achieve a clearer baseline for comparisons, four structures were eliminated, one from each condition: Adjective placement (Double mismatch condition), addition of a definite article prior to mass nouns (L1 mismatch condition), past progressive tense (L2 mismatch condition), and tense sequence (control condition). After eliminating these structures, accuracy of the native English speakers was equated across conditions [ $F(3,108) = 0.82$ ,  $MSE = 0.025$ ,  $p = 0.48$ ,  $\eta_p^2 = 0.022$ ]. The final set therefore included 9 constructions: 3 in the Double mismatch condition (11 items), 2 in the L1 mismatch condition (10 items), 2 in the L2 mismatch condition (10 items), and 2 in the control condition (10 items). Performance of the native English speakers on these remaining 9 constructions (see **Table 3**) was considered the baseline of performance in online and offline measures against which the performance of the AHE trilinguals was examined.

### Analyses Approach and Model Structure

Reading measures were analyzed for the target word in each sentence, defined as the point in the sentence at which the grammatical violation became apparent. Thus, in the case where an obligatory constituent was omitted from the sentence to create the violation the following word was defined as the target word (in the sentence “I am planning to buy \*dog for my son’s tenth birthday,” the word “dog” was defined as the target word). In

**TABLE 3 |** List of remaining syntactic categories with different degrees of overlap in Arabic, Hebrew, and English.

| Cross language influence conditions |                                                                     |                                                                       |                                         |                                                           |
|-------------------------------------|---------------------------------------------------------------------|-----------------------------------------------------------------------|-----------------------------------------|-----------------------------------------------------------|
|                                     | Double mismatch (interference from both L1 and L2)                  | L1 mismatch (interference only from L1)                               | L2 mismatch (interference only from L2) | Control condition (no interference)                       |
| Structures                          | First person pro-drop; Copula omission; Indefinite article omission | Possessive marking; Definite article omission in the superlative form | Comparative form; Superlative form      | Verb-time expression agreement; Quantifier-noun agreement |

cases where an incorrect form was used, it was defined as the target word (in the sentence “Everyone knows that I’m the \*most rich in this neighborhood” the word “rich” was defined as the target word).

Grammaticality judgment accuracy, as well as 6 measures from the eye tracking record (First Fixation Duration, Gaze Duration, Total Time, Skipping, Regressions Out, Regressions In) were analyzed using linear mixed effect models, as these models allow one to simultaneously account for variance related to participants and to items. Grammaticality Judgments, Skipping and Regressions (In and Out) were analyzed following a binomial distribution (i.e., mixed logistic regression). Duration measures were log transformed to reduce skew in the distribution, as these transformations improved normality more than the inverse transformation. Within each measure, we first identified significant control variables, which were retained in subsequent models. Specifically, we used the *buildmer* function in the *buildmer* package (v. 1.3, Voeten, 2019) in R (version 4.0.3, R Core Team, 2020), which uses the (g)lmer function from the *lme4* package (v 1.1.1-21, Bates et al., 2015), to fit a model including all (normalized) control variables (participants’ age, target length, target frequency, sentence length in characters, average frequency of the words in the sentence, and averaged length of the words in the sentence). Using backward stepwise elimination, the *buildmer* function calculates p-values for all fixed effects based on Satterthwaite degrees of freedom using the *lmerTest* package (v. 3.1-0, Kuznetsova et al., 2017), or the Wald degrees of freedom for binomial distribution.

Once control variables were identified for each measure, we compared an additive model including the effects of Group and Mismatch Type (Model 1) against an interactive model including in addition the interaction between Group and Mismatch Type (Model 2) using Log Likelihood Ratio Test. The factors of interest were coded using treatment/dummy coding, such that for the effect of Group, Arabic-Hebrew-English (AHE) trilinguals were set as the reference against which native English (NE) speakers were compared. Similarly, for the effect of Mismatch Type, Control sentences were set as the reference against which L1 Mismatch, L2 Mismatch, and Double Mismatch sentences were compared. The random structure included by-participant and by-item intercepts, as well as by-participant slope for Mismatch Type and by-item slope for Group. In case of convergence failure, the random structure was simplified following the guidelines provided by Poort and Rodd (2019, removing correlations, removing slope with lowest variance while reintroducing correlations; removing correlations; removing the other slope). To probe interactive effects and conduct pairwise comparisons, we used the *testInteractions* function from the *phia* package (v.0.2-1, De Rosario-Martinez, 2015) with Bonferroni adjustments for multiple comparisons. Estimated means and standard errors (SE) were obtained via the *emmeans* package (v.1.5. 2-1, Lenth, 2020) the full R script of the analyses can be found in the **Supplementary Materials**.

## Analyses

**Table 4** provides observed mean performance for each measure as a function of Group and Mismatch Type.

**Grammaticality Judgment.** In the Grammaticality Judgment measure, model comparisons revealed that the interactive model (M2) improved the fit over the additive model [ $\chi^2(3) = 13.75, p = 0.003$ ]. Examination of model summary (see **Table 6**) revealed an interaction between Group and the difference between L2 Mismatch and Control, as well as a marginal interaction of Group with the difference between L1 Mismatch and Control. As seen in **Figure 1**, and supported by the pairwise comparisons with Bonferroni corrections for multiple comparisons (**Table 7**), the difference between L2 Mismatch and Control was larger for AHE [ $b = 0.80, \chi^2(1) = 12.93, p = 0.004$ ] than for NE [ $b = 0.34, \chi^2(1) = 1.82, p = 1.00$ ]. Further, the difference between L1 Mismatch and control was marginally significant for AHE [ $b = 0.71, \chi^2(1) = 7.57, p = 0.071$ ] but not for NE [ $b = 0.50, \chi^2(1) = 0.00, p = 1.00$ ]. Recall that the accuracy levels of the NE in the grammaticality judgment task was used to select the subset of constructions on which to examine the performance of AHE. Thus, it is not surprising that there are no differences across conditions in the NE group.

**First Fixation Duration.** In the FFD measure, model comparisons revealed that the interactive model (M2) did not improve the fit over the additive model [ $\chi^2(3) = 4.53, p = 0.21$ ]. Examination of model summary (**Table 5**) revealed a significant effect of Group, such that NE speakers had shorter FFD ( $M = 239, 95\% \text{ CI } [229, 249]$ ) compared to AHE ( $M = 260, 95\% \text{ CI } [250, 271]$ ).

**Gaze Duration.** In the GD measure, model comparisons revealed that the interactive model (M2) improved the fit over the additive model [ $\chi^2(3) = 8.74, p = 0.03$ ]. Examination of model summary (see **Table 5**) revealed that the difference between L1 Mismatch sentences and Controls was modulated by Group (see **Figure 2**). Pairwise comparisons with Bonferroni corrections for multiple comparisons (**Table 7**) revealed however, that only the difference between L1 Mismatch and L2 Mismatch in the AHE reached significance [ $b = 0.28, \chi^2(1) = 10.90, p = 0.011$ ].

**Total Time.** In the TT measure, model comparisons revealed that the interactive model (M2) did not improve the fit over the additive model [ $\chi^2(3) = 1.58, p = 0.66$ ]. Examination of model summary (**Table 5**) revealed a significant effect of Group, such that NE speakers had shorter Total reading times ( $M = 330, 95\% \text{ CI } [294, 371]$ ) compared to AHE ( $M = 827, 95\% \text{ CI } [739, 927]$ ). In addition, total reading times for the target word in the L2 Mismatch condition were shorter ( $M = 476, 95\% \text{ CI } [426, 533]$ ) than for target words in the control No Mismatch condition ( $M = 593, 95\% \text{ CI } [531, 662]$ ).

**Skipping.** When examining Skipping Rates, model comparisons revealed that the interactive model (M2) did not improve the fit over the additive model [ $\chi^2(3) = 0.43, p = 0.93$ ]. Examination of model summary (**Table 6**) revealed an effect of Group, such that NE speakers skipped the target word more often ( $M = 0.15, 95\% \text{ CI } [0.12, 0.19]$ ) than AHE trilinguals ( $M = 0.02, 95\% \text{ CI } [0.01, 0.03]$ ).

**Regressions Out.** For the Regressions Out, model comparisons revealed that the interactive model (M2) did not improve the fit over the additive model [ $\chi^2(3) = 4.69, p = 0.20$ ]. Examination of model summary (**Table 6**) revealed more regression out of targets in the L1 Mismatch condition ( $M = 0.30, 95\% \text{ CI } [0.23, 0.37]$ ).

**TABLE 4 |** Observed mean performance (SE) as a function of Group and Mismatch Type.

| Group | Mismatch type   | Measure        |         |          |           |               |                 |                |
|-------|-----------------|----------------|---------|----------|-----------|---------------|-----------------|----------------|
|       |                 | Gram. judgment | FFD     | GD       | TT        | Skipping rate | Regressions out | Regressions in |
| AHE   | Control         | 0.62 (0.03)    | 288 (7) | 499 (20) | 1058 (43) | 0.03 (0.01)   | 0.26 (0.02)     | 0.56 (0.02)    |
|       | L1 Mismatch     | 0.44 (0.02)    | 281 (7) | 600 (24) | 1285 (52) | 0.02 (0.01)   | 0.32 (0.02)     | 0.53 (0.02)    |
|       | L2 Mismatch     | 0.36 (0.02)    | 291 (8) | 394 (14) | 883 (40)  | 0.05 (0.01)   | 0.19 (0.02)     | 0.54 (0.02)    |
|       | Double Mismatch | 0.56 (0.02)    | 286 (7) | 479 (19) | 1088 (41) | 0.04 (0.01)   | 0.26 (0.02)     | 0.61 (0.02)    |
| NE    | Control         | 0.81 (0.02)    | 263 (6) | 307 (10) | 431 (15)  | 0.21 (0.02)   | 0.18 (0.03)     | 0.36 (0.03)    |
|       | L1 Mismatch     | 0.79 (0.02)    | 248 (5) | 288 (9)  | 445 (15)  | 0.13 (0.02)   | 0.31 (0.03)     | 0.39 (0.03)    |
|       | L2 Mismatch     | 0.83 (0.02)    | 239 (5) | 257 (7)  | 321 (12)  | 0.24 (0.02)   | 0.22 (0.03)     | 0.24 (0.03)    |
|       | Double Mismatch | 0.77 (0.02)    | 259 (6) | 302 (9)  | 424 (13)  | 0.21 (0.02)   | 0.18 (0.02)     | 0.40 (0.03)    |

FFD, first fixation durations; GD, gaze durations; TT, total reading times.

SE calculated over all data points taking into account the presence of within-participant variables following (Morey, 2008). To this end, we used the function described by Change, W. [http://www.cookbook-r.com/Graphs/Plotting\\_means\\_and\\_error\\_bars\\_\(ggplot2\)](http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2))

**TABLE 5 |** LME models predicting reading times (FFD, GD, TT).

| Fixed effects                | FFD      |      |           | GD    |          |          | TT    |      |           |
|------------------------------|----------|------|-----------|-------|----------|----------|-------|------|-----------|
|                              | b        | SE   | t         | b     | SE       | t        | b     | SE   | t         |
| (intercept)                  | 5.58     | 0.03 | 196.64*** | 5.98  | 0.07     | 85.70*** | 6.78  | 0.07 | 96.16***  |
| Group (NE)                   | −0.09    | 0.03 | −2.99**   | −0.34 | 0.10     | −3.60*** | −0.92 | 0.09 | −10.53*** |
| Mismatch (L1)                | −0.02    | 0.03 | −0.54     | 0.16  | 0.09     | 1.90*    | 0.00  | 0.06 | −0.07     |
| Mismatch (L2)                | −0.04    | 0.03 | −1.45     | −0.12 | 0.09     | −1.40    | −0.22 | 0.07 | −3.35**   |
| Mismatch (Double)            | 0.00     | 0.03 | 0.08      | −0.04 | 0.08     | −0.45    | −0.04 | 0.06 | −0.61     |
| Group(NE):Mismatch(L1)       | −        | −    | −         | −0.27 | 0.11     | −2.35*   | −     | −    | −         |
| Group(NE): Mismatch (L2)     | −        | −    | −         | 0.01  | 0.12     | 0.06     | −     | −    | −         |
| Group(NE): Mismatch (Double) | −        | −    | −         | 0.01  | 0.11     | 0.08     | −     | −    | −         |
| <b>Control variables</b>     |          |      |           |       |          |          |       |      |           |
| Participant's age            | −        | −    | −         | −     | −        | −        | −0.08 | 0.04 | −2.17*    |
| Target length                | −0.03    | 0.01 | −2.79**   | 0.07  | 0.01     | 5.08***  | 0.10  | 0.02 | 4.11***   |
| Target frequency             | −        | −    | −         | −     | −        | −        | −0.08 | 0.02 | −3.39**   |
| Random effects               | Variance |      | SD        |       | Variance |          | SD    |      |           |
|                              |          |      |           |       |          |          |       |      |           |
| Item (intercept)             | 0.00     | 0.05 | 0.03      | 0.17  | 0.04     | 0.20     |       |      |           |
| Group (NE)                   | 0.00     | 0.04 | 0.00      | 0.06  | 0.06     | 0.25     |       |      |           |
| Group (AHE)                  | −        | −    | 0.03      | 0.17  | −        | −        |       |      |           |
| Participant (intercept)      | 0.02     | 0.15 | 0.00      | 0.00  | 0.11     | 0.33     |       |      |           |
| Mismatch Type (L1)           | 0.01     | 0.12 | 0.04      | 0.21  | 0.02     | 0.12     |       |      |           |
| Mismatch Type (L2)           | 0.02     | 0.15 | 0.03      | 0.16  | 0.05     | 0.22     |       |      |           |
| Mismatch Type (Double)       | 0.01     | 0.07 | 0.04      | 0.19  | 0.02     | 0.13     |       |      |           |
| Mismatch Type (Control)      | −        | −    | 0.06      | 0.24  | −        | −        |       |      |           |
| Residual                     | 0.13     | 0.36 | 0.22      | 0.47  | 0.30     | 0.55     |       |      |           |

Effect sizes (bs), standard errors (SEs), and t-values.

FFD, first fixation durations; GD, gaze durations; TT, total reading times.

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ ; •marginal.

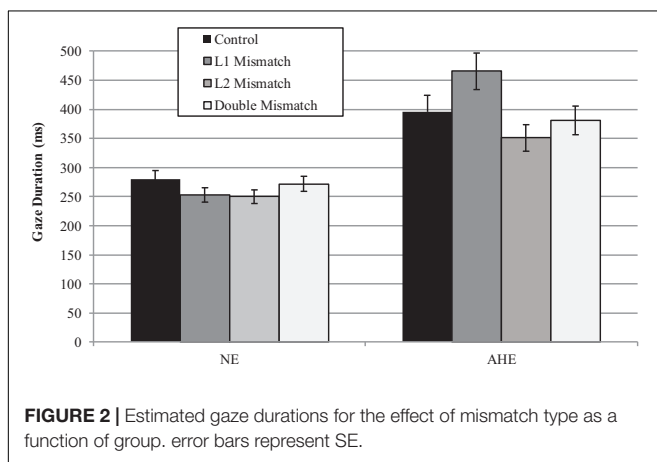
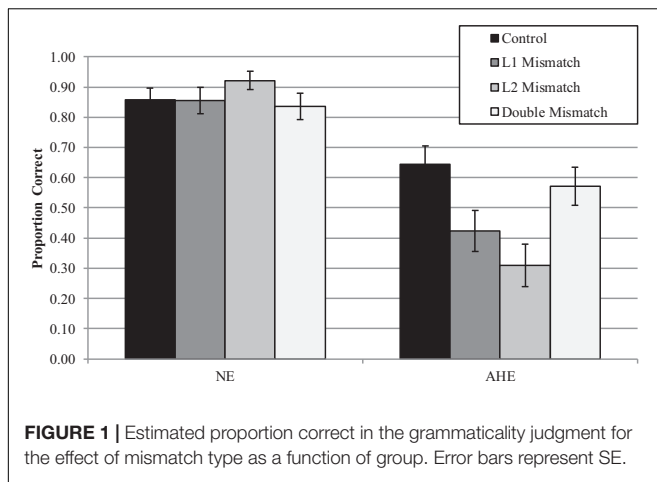
relative to Control ( $M = 0.20$ , 95% CI [0.15, 0.26]), but this effect was not modulated by Group.

**Regression In.** When examining Regression Into the target area, model comparisons revealed that the interactive model (M2) improved the fit over the additive model [ $\chi^2(3) = 9.05$ ,  $p = 0.03$ ]. Examination of model summary (Table 6) revealed that the difference between L2 Mismatch sentences and Controls was modulated by Group. As seen in Figure 3, and supported by pairwise comparisons with Bonferroni corrections (Table 7),

whereas NE control participants regressed less into targets of sentences in the L2 Mismatch condition relative to the other conditions, this difference was not present for AHE trilinguals.

## DISCUSSION

The current study examined L1 and L2 as potential sources of cross-language influences during L3 processing when typological



similarity is neutralized. In accordance with our hypotheses, we observed interference from both background languages in L3 processing. L1 interference was observed in earlier online measures of processing than L2 interference, and both were also observed in offline judgments. Surprisingly, however, whereas structural mismatch with a single background language (either the L1 or the L2) resulted in significant interference, structural mismatch with both background languages did not lead to significant interference. We address each of these findings below.

The current findings demonstrate that structural mismatch between the L3 and either the L1 or the L2 of trilingual speakers resulted in significant interference. Specifically, participants were less accurate at identifying ungrammatical sentences in English when the corresponding structure was grammatical in the L2, and marginally so when the structure was grammatical in the L1. Further, participants had longer gaze durations in the critical target area when reading L1 mismatch ungrammatical sentences. When reading L2 mismatch ungrammatical sentences, first pass reading times were not affected but participants made more regressions back into the target area than would be expected based on the performance of native English speakers who do not experience cross-language influence. These findings indicate the presence of interference from both the L1 and the L2 of trilingual speakers, and align with the theoretical stance put

forth by the Linguistic Proximity Model (Westergaard et al., 2017) and Scalpel Model (Slabakova, 2017). According to these models, cross-language influence is determined on a structure by structure basis, and neither background language has a privileged role in supplying cross-language influences.

Extending previous literature, the current study further reveals time-course differences in the operation of cross-language influence from L1 and L2. Specifically, interference from L1 (Arabic) was evident early on, in first pass reading measures. Gaze durations to target words in L1 mismatch structures were longer than gaze durations to target words in sentences with control structures, for AHE trilinguals but not for monolingual English speakers. Sensitivity to L1 interference was not apparent in later reading measures, such as total time and regressions in, but was marginally significant in the offline grammaticality judgment measure. Interference from L2 (Hebrew), however, was not evident in the early measures of online processing (first fixation durations, gaze durations, skipping or regression out), but was apparent in the second pass reading measure of regressions into the target region. In comparison to the monolingual English speakers, who exhibited reduced regressions into the target region in sentences in the L2 mismatch condition, the AHE trilinguals exhibited equal rates of regressions into the target region in the L2 mismatch and Control conditions. We interpret this pattern as indicating that the structures in the L2 mismatch condition were easier for native English speakers than those in the Control condition, given that these are different structures, but critically that AHE trilinguals did not show this expected facilitation due to interference from the L2. The offline measure supports this interpretation, as the AHE trilinguals were much less accurate in their grammaticality judgment decisions on sentences in the L2 mismatch condition compared to control sentences.

Thus, the answer to the question which of a trilinguals' background languages exerts stronger cross-language influence during L3 processing appears to depend on the measure. Specifically, L1 interference was evident earlier in processing, but L2 interference was stronger in the offline metalinguistic measure. One possible explanation for this pattern is that because trilinguals were sensitive to L1 interference already during first pass reading, they were more successful in resolving this interference by the time they performed the grammaticality judgment after completing reading the sentence. In contrast, because sensitivity to L2 interference emerged only later in sentence processing, in second pass reading measures, it was not yet resolved, and thus exerted a stronger influence on sentence final grammaticality judgments.

By adding sensitive measures of cross-language influence during online processing, we were able to identify a nuanced pattern of results. Specifically, although cross-language influence from the L1 was only marginal in the sentence-final grammaticality judgment, it was robust during the earlier measure of reading time. Further, the difference in timing between cross-language influences from the L1 and the L2 only emerged in the real-time online measures. Such divergence between real-time online measures and offline grammaticality judgments has been observed and influential in previous studies on L2 learning (e.g., McLaughlin et al., 2004; Tokowicz and

**TABLE 6 |** LME models predicting grammaticality judgment, skipping rate, regressions in, and regressions out.

| Fixed effects                | Gram judgment |      |          | Skipping rate |          |           | Regression in |      |          | Regression out |      |          |
|------------------------------|---------------|------|----------|---------------|----------|-----------|---------------|------|----------|----------------|------|----------|
|                              | b             | SE   | z        | b             | SE       | z         | b             | SE   | z        | b              | SE   | z        |
| (intercept)                  | 0.59          | 0.26 | 2.25*    | −3.78         | 0.27     | −13.80*** | 0.30          | 0.18 | 1.61     | −1.29          | 0.20 | −6.43*** |
| Group (NE)                   | 1.19          | 0.39 | 3.04**   | 2.07          | 0.25     | 8.23***   | −0.94         | 0.27 | −3.37*** | −0.24          | 0.21 | −1.12    |
| Mismatch (L1)                | −0.89         | 0.32 | −2.75**  | −0.32         | 0.30     | −1.05     | −0.17         | 0.18 | −0.93    | 0.52           | 0.22 | 2.32*    |
| Mismatch (L2)                | −1.39         | 0.38 | −3.59*** | 0.03          | 0.24     | 0.16      | −0.09         | 0.19 | −0.48    | −0.30          | 0.24 | −1.22    |
| Mismatch (Double)            | −0.30         | 0.32 | −0.92    | 0.17          | 0.25     | 0.69      | 0.23          | 0.18 | 1.30     | −0.03          | 0.22 | −0.16    |
| Group(NE): Mismatch (L1)     | 0.87          | 0.47 | 1.84*    | –             | –        | –         | 0.31          | 0.27 | 1.16     | –              | –    | –        |
| Group(NE): Mismatch (L2)     | 2.05          | 0.58 | 3.50***  | –             | –        | –         | −0.6          | 0.30 | −1.98*   | –              | –    | –        |
| Group(NE): Mismatch (Double) | 0.13          | 0.47 | 0.28     | –             | –        | –         | −0.06         | 0.26 | −0.23    | –              | –    | –        |
| <b>Control variables</b>     |               |      |          |               |          |           |               |      |          |                |      |          |
| Participant's Age            | 0.06          | 0.12 | 0.55     | −0.12         | 0.10     | −1.2      | –             | –    | –        | –              | –    | –        |
| Target Length                | –             | –    | –        | −0.67         | 0.10     | −6.54***  | –             | –    | –        | –              | –    | –        |
| Target Frequency             | –             | –    | –        | –             | –        | –         | –             | –    | –        | –              | –    | –        |
| Random effects               | Variance      | SD   | Variance | SD            | Variance | SD        | Variance      | SD   |          |                |      |          |
|                              |               |      |          |               |          |           |               |      |          |                |      |          |
| Item (intercept)             | 0.36          | 0.60 | 0.36     | 0.60          | 0.05     | 0.23      | 0.19          | 0.44 |          |                |      |          |
| Group (NE)                   | 0.69          | 0.83 | 0.29     | 0.54          | 0.10     | 0.32      | 0.44          | 0.66 |          |                |      |          |
| Group (AHE)                  | –             | –    | –        | –             | –        | –         | –             | –    |          |                |      |          |
| Participant (intercept)      | 0.74          | 0.86 | 0.33     | 0.58          | 0.81     | 0.90      | 0.47          | 0.69 |          |                |      |          |
| Mismatch Type (L1)           | 0.38          | 0.62 | 0.43     | 0.65          | 0.00     | 0.08      | 0.12          | 0.35 |          |                |      |          |
| Mismatch Type (L2)           | 2.16          | 1.47 | 0.05     | 0.22          | 0.20     | 0.44      | 0.32          | 0.56 |          |                |      |          |
| Mismatch Type (Double)       | 0.63          | 0.79 | 0.21     | 0.45          | 0.00     | 0.07      | 0.01          | 0.11 |          |                |      |          |
| Mismatch Type (Control)      | –             | –    | –        | –             | –        | –         | –             | –    |          |                |      |          |

Effect sizes (bs), standard errors (SEs), and z-values.

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ ; •marginal.

MacWhinney, 2005). We therefore believe that incorporating similar measures of online processing to studies of cross-language influences in L3 processing is a fruitful avenue which might be useful in reconciling some of the conflicting findings in the extant literature. Nevertheless, it is important to note that our experimental design incorporated both eye-movement data and grammaticality judgments following each sentence, which could have influenced natural reading and activated greater metalinguistic awareness (Valdés Kroff et al., 2018).

Therefore, future research should continue to investigate this issue by separating online and offline tasks in order to maintain cleaner measures.

We predicted the strongest degree of interference in structures that mismatch both the L1 and the L2, because the entirety of participants' background linguistic knowledge conflicted with the L3 in these structures. However, not only did we not find stronger interference under these conditions, but in fact interference was not significant for these structures across the different measures – either online or offline. We propose that these structures may have been particularly salient for the AHE trilinguals tested in the current study, for one of two possible reasons. First, it is possible that when an L3 learner encounters a structure that differs from both of her background languages this draws attention and emphasizes the need to relearn a morpho-syntactic feature of the L3 (Schmidt, 2012). Thus, the morpho-syntactic structures in this category can be considered as being unique to the trilinguals' L3, because they are unattested in either the L1 or the L2, and may thus gain particular salience to learners (Tokowicz and MacWhinney, 2005; Tolentino and Tokowicz, 2014). As a result, speakers may become more aware of the potential error on such structures and thus monitor their performance on this feature more closely. Second, it is possible that these specific morpho-syntactic features of English as an L3 are explicitly highlighted during instruction because of the mismatch with participants' background languages. These options are

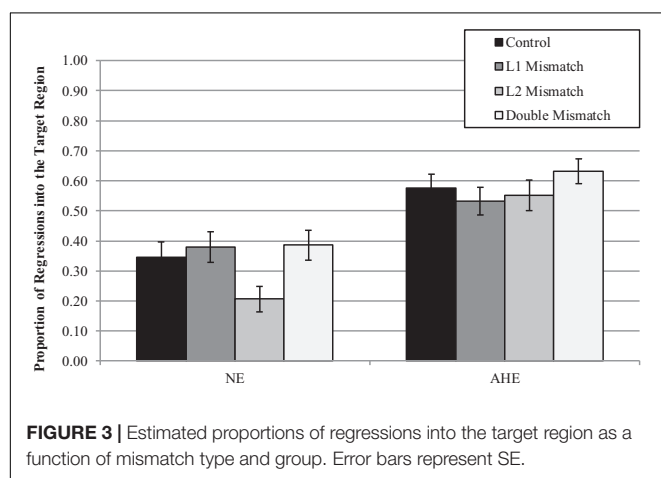


**TABLE 7 |** Summary of pairwise comparisons with Bonferroni corrections for multiple comparisons as a function of Mismatch Type and Group.

| Comparison |                         | Grammaticality Judgment |          |          | Gaze Duration |          |          | Regression In |          |          |
|------------|-------------------------|-------------------------|----------|----------|---------------|----------|----------|---------------|----------|----------|
|            |                         | <i>b</i>                | $\chi^2$ | <i>p</i> | <i>b</i>      | $\chi^2$ | <i>p</i> | <i>b</i>      | $\chi^2$ | <i>p</i> |
| AHE        | Control vs. L1 Mis.     | 0.71                    | 7.57     | 0.07     | −0.16         | 3.60     | 0.69     | 0.54          | 0.88     | 1.00     |
|            | Control vs. L2 Mis.     | 0.80                    | 12.93    | 0.004**  | 0.12          | 1.95     | 1.00     | 0.52          | 0.24     | 1.00     |
|            | Control vs. Double Mis. | 0.58                    | 0.85     | 1.00     | 0.04          | 0.20     | 1.00     | 0.44          | 1.71     | 1.00     |
|            | L1 Mis. vs. L2 Mis.     | 0.62                    | 1.71     | 1.00     | 0.28          | 10.90    | 0.012*   | 0.48          | 0.15     | 1.00     |
|            | L1 Mis. vs. Double Mis. | 0.36                    | 3.68     | 0.66     | 0.20          | 5.99     | 0.17     | 0.40          | 5.18     | 0.27     |
|            | L2 Mis. vs. Double Mis. | 0.75                    | 7.61     | 0.07*    | 0.08          | 1.00     | 1.00     | 0.58          | 3.01     | 1.00     |
| NE         | Control vs. L1 Mis.     | 0.50                    | 0.00     | 1.00     | 0.10          | 3.78     | 0.62     | 0.46          | 0.37     | 1.00     |
|            | Control vs. L2 Mis.     | 0.34                    | 1.82     | 1.00     | 0.11          | 4.26     | 0.47     | 0.67          | 7.05     | 0.095*   |
|            | Control vs. Double Mis. | 0.54                    | 0.15     | 1.00     | 0.03          | 0.31     | 1.00     | 0.46          | 0.56     | 1.00     |
|            | L1 Mis. vs. L2 Mis.     | 0.34                    | 1.95     | 1.00     | 0.01          | 0.03     | 1.00     | 0.70          | 10.72    | 0.01*    |
|            | L1 Mis. vs. Double Mis. | 0.54                    | 0.14     | 1.00     | −0.08         | 2.40     | 1.00     | 0.49          | 0.02     | 1.00     |
|            | L2 Mis. vs. Double Mis. | 0.30                    | 2.76     | 1.00     | 0.08          | 2.71     | 1.00     | 0.71          | 11.74    | 0.007**  |

\*\*\**p* < 0.001; \*\**p* < 0.01; \**p* < 0.05; •*marginal*.

not mutually exclusive, but future research might be able to distinguish among them by testing less proficient trilinguals from the same population. If the driving force is explicit instruction, less proficient trilinguals would demonstrate relative ease with processing such Double mismatch structures just as high proficiency trilinguals do. If, however, this facility in processing arises slowly with growing L3 proficiency and meta linguistic knowledge, we would expect lower proficiency trilinguals to indeed show increased interference for the Double mismatch structures.

Extending previous studies of L3 learning, which have largely focused on the initial stages and on individuals who have self-selected to become multilinguals (e.g., Ecke and Hall, 2013), here we tested individuals who have become trilinguals due to their socio-educational context and are moderately proficient users of L3 English. Thus, the results of the current study carry the potential to be more generalizable to typical multilinguals in today's global society (Kaushanskaya and Prior, 2015).

Further, our approach to selecting language stimuli differs from that adopted by most previous research. Namely, we identified a wider number of syntactic structures different from each other in their mismatch with participants' L1 and L2. This allowed us to simultaneously measure cross-language influences from both background languages in a single group of participants, which has the advantage of greatly reducing potential differences (in language learning background, proficiency in L1/L2) that might arise in between-participant comparisons, even in "Mirror Group" designs (Giancaspro et al., 2015). However, this approach has the inevitable result that the experimental conditions included different syntactic structures, which introduces a different source of variability, such as potential baseline differences in sensitivity or salience of the selected structures. Even though some structures were eliminated so that accuracy of the native English speaking control group in the grammaticality judgment was equated across conditions, it is possible that

some variability remained unaccounted for. Indeed, the pattern observed in the regressions-in measure for native English monolingual speakers suggests such baseline variability. Thus, the current results should be interpreted as complementary to those arising from other methodologies, to lead to a fuller nuanced understanding. Future research can explore alternative means of matching between structures, or directly compare the results of experiments using these different methodological approaches.

Finally, cross-language influences can manifest as either facilitation or interference, the latter of which was the focus of the current study. Importantly, it is currently unclear whether facilitation and interference effects are symmetrical, and whether they are similarly easy to detect. Indeed, often the direction of influence is determined by the type of manipulation examined in a particular study, and specifically by how researchers define the baseline condition. Thus, in the lexical domain for instance, items that are non-cognates typically serve as controls, against which overlap in form and meaning (cognates) results in facilitation but overlap in form (but not meaning) results in interference (at least in processing, but not in learning, e.g., Hirosh and Degani, (accepted); Elias and Degani, unpublished; Marecka et al., 2020). In the syntactic domain, some researchers have treated unique syntactic constructions as a baseline, such that when constructions that are similar across languages are compared to this baseline facilitation is expected, but when structures that differ across languages are compared to the unique baseline condition, interference is expected (e.g., Tokowicz and MacWhinney, 2005). The pattern of results, however, is more complex, as unique structures are sometimes experienced as especially difficult (Tokowicz and Warren, 2010). Somewhat analogously, in the current study we defined our baseline as constructions that are shared across all three languages, such that constructions that are not shared by (at least) one language will index interference. Our matching procedures were therefore conducted between the interference conditions and the selected baseline. Alternatively, one could have selected constructions

unique to English as the baseline, against which constructions that also overlap with the L1, with the L2 or with both, will index facilitation. Future studies may be useful in examining whether facilitative and interfering cross-language influences operate similarly for bilingual and trilingual speakers.

## CONCLUSION

Overall, the findings of the current study suggest that the entire linguistic repertoire is activated when processing L3. These findings are consistent with the Linguistic Proximity Model (Westergaard et al., 2017) and the Scalpel model (Slabakova, 2017). Our results demonstrate that neither L1 nor L2 are the single default supplier for cross-language influence, and that all previously learned languages may shift and modulate the linguistic system. Going beyond typological proximity, the current study documents robust cross-language influences across languages that are typologically distinct. Finally, the current study sheds light on the difference between performance in online and offline measures, and how processing language in real-time differs from judgments that rely on meta-linguistic knowledge (Dussias, 2010; Sedivy, 2010). In addition to theoretical insights, the current study has important implications for L3 language instruction. Specifically, our results suggest that difficulties in L3 learning might not only be a result of interference from the L1 (Tajareh, 2015), but could also reflect cross-language influences from the L2. Thus, when scaffolding L3 learning, both L1 and L2 should be taken into account as influential background languages.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## REFERENCES

- Angelovska, T., and Hahn, A. (2012). "Written L3 (English): transfer phenomena of L2 (German) lexical and syntactic properties," in *Cross-linguistic Influences in Multilingual Language Acquisition*, ed. D. Gabrys-Barker, (Berlin: Springer), 23–40. doi: 10.1007/978-3-642-29557-7\_2
- Bardel, C., and Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Sec. Lang. Res.* 23, 459–484. doi: 10.1177/0267658307080557
- Bardel, C., and Falk, Y. (2012). "Behind the L2 status factor: a neurolinguistic framework for L3 research," in *Third Language Acquisition in Adulthood*, eds J. Cabrelli Amaro, and J. Rothman, (Amsterdam: John Benjamins), 61–78.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bruhn de Garavito, J., and Perpiñán, S. (2014). "Subject pronouns and clitics in the Spanish interlanguage of French L1 speakers," in *Proceedings of the 2014 Annual Conference of the Canadian Linguistic Association*, ed. L. Teddman (Ottawa: Canadian Linguistic Association), 1–11.
- Cabrelli-Amaro, J., Amaro, J. F., and Rothman, J. (2015). "The relationship between L3 transfer and structural similarity across developments: raising across

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Education at the University of Haifa. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

NA, TD, and AP designed the study, analyzed data, and wrote the manuscript. NA created stimuli and collected data. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by ISF grant 340/18 to AP and TD. NA was supported by funding from the Arabic Language Academy, and by Edmond J. Safra Brain Research Center, the Jewish-Arabic Center, and the Graduate Studies Authority, all at the University of Haifa.

## ACKNOWLEDGMENTS

The authors thank Margarita Kaushanskaya for her assistance in data collection from native English speakers in her lab, Razan Silawi for assistance in data collection and coding, and Nur Awwad for her work on stimuli development.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://osf.io/c625n/> (identifier: DOI 10.17605/OSF.IO/C625N).

- an experiences in Brazilian Portuguese," in *Transfer Effects in Multilingual Language Development*, ed. H. Peukert, (Amsterdam: John Benjamins), 21–52. doi: 10.1075/hsld.4.02ama
- Cenoz, J. (2001). "The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition," in *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, Vol. 31, eds J. Cenoz, B. Hufeisen, and U. Jessner, (Clevedon: Multilingual Matters), 8–20. doi: 10.21832/9781853595509-002
- Conklin, K., and Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Sec. Lang. Res.* 32, 453–467. doi: 10.1177/0267658316637401
- De Angelis, G., and Selinker, L. (2001). "Interlanguage transfer and competing linguistic systems in the multilingual mind," in *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, Vol. 31, eds J. Cenoz, B. Hufeisen, and U. Jessner, (Clevedon: Multilingual Matters), 42–58. doi: 10.21832/9781853595509-004
- Degani, T., Prior, A., and Tokowicz, N. (2011). Bidirectional transfer: the effect of sharing a translation. *J. Cogn. Psychol.* 23, 18–28. doi: 10.1080/20445911.2011.445986
- De Rosario-Martinez, H. (2015). *Analyzing Interactions of Fitted Models*. CRAN-R Project.

- Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annu. Rev. Appl. Linguist.* 30, 149–166. doi: 10.1017/S026719051000005X
- Dussias, P. E., and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish English bilinguals. *Bilingual Lang. Cogn.* 10, 101–116. doi: 10.1017/S1366728906002847
- Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingual Lang. Cogn.* 18, 145–162. doi: 10.1017/S1366728913000722
- Ecke, P., and Hall, C. J. (2013). Tracking tip-of-the-tongue states in a multilingual speaker: evidence of attrition or instability in lexical systems? *Int. J. Bilingual.* 17, 734–751. doi: 10.1177/1367006912454623
- Falk, Y., and Bardel, C. (2010). The study of the role of the background languages in third language acquisition. The state of the art. *IRAL Int. Rev. Appl. Linguist. Lang. Teach.* 48, 185–219. doi: 10.1515/iral.2010.009
- Falk, Y., and Bardel, C. (2011). Object pronouns in German L3 syntax: evidence for the L2 status factor. *Sec. Lang. Res.* 27, 59–82. doi: 10.1177/0267658310386647
- Fallah, N., Jabbari, A. A., and Fazilatfar, A. M. (2016). Source (s) of syntactic cross linguistic influence (CLI): the case of L3 acquisition of english possessives by Mazandarani Persian bilinguals. *Sec. Lang. Res.* 32, 225–245. doi: 10.1177/0267658315618009
- Flynn, S., Foley, C., and Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *Int. J. Multilingual.* 1, 3–16. doi: 10.1080/14790710408668175
- Giancaspro, D., Halloran, B., and Iverson, M. (2015). Transfer at the initial stages of L3 Brazilian portuguese: a look at three groups of english/spanish bilinguals. *Bilingual Lang. Cogn.* 18, 191–207. doi: 10.1017/S1366728914000339
- Godfroid, A., and Winke, P. (2015). "Investigating implicit and explicit processing using L2 learners' eye-movement data," in *Implicit and Explicit Learning of Languages*, ed. P. Rebuschat, (Amsterdam: John Benjamins), 325–348. doi: 10.1075/sibil.48.14god
- Gollan, T. H., Montoya, R. I., and Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16, 562–576. doi: 10.1037/0894-4105.16.4.562
- Hammarberg, B. (2001). "Roles of L1 and L2 in L3 production and acquisition," in *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, Vol. 31, eds J. Cenoz, B. Hufeisen, and U. Jessner, (Clevedon: Multilingual Matters), 21–41. doi: 10.21832/9781853595509-003
- Hammarberg, B. (2010). The languages of the multilingual: some conceptual and terminological issues. *IRAL Int. Rev. Appl. Linguist. Lang. Teach.* 48, 91–104. doi: 10.1515/iral.2010.005
- Hermas, A. (2010). Language acquisition as computational resetting: verb movement in L3 initial state. *Int. J. Multilingual.* 7, 343–362. doi: 10.1080/14790718.2010.487941
- Hirosh, Z., Degani, T. (accepted) Bilingual novel word learning as a function of language of instruction: is learning better through L1 or L2? *Lang. Learn.* 1–34.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: performance similarities between non-native and native speakers. *Lingua* 120, 901–931. doi: 10.1016/j.lingua.2009.06.004
- Hopp, H. (2019). Cross-linguistic influence in the child third language acquisition of grammar: Sentence comprehension and production among Turkish-German and German learners of English. *Int. J. Bilingual.* 23, 567–583. doi: 10.1177/1367006917752523
- Isurin, L. (2005). "Cross-linguistic transfer in word order: evidence from L1 forgetting and L2 acquisition," in *Proceedings of the 4th International Symposium on Bilingualism*, eds J. Cohen, K. T. McAlister, K. Rolstad, and J. MacSwan, (Somerville, MA: Cascadilla Press), 1115–1130.
- Kaushanskaya, M., and Prior, A. (2015). Variability in the effects of bilingualism on cognition: it is not just about cognition, it is also about bilingualism. *Bilingual Lang. Cogn.* 18, 27–28. doi: 10.1017/S1366728914000510
- Kavé, G. (2005). Phonemic fluency, semantic fluency, and difference scores: normative data for adult Hebrew speakers. *J. Clin. Exp. Neuropsychol.* 27, 690–699. doi: 10.1080/13803390490918499
- Keating, G. D., and Jegerski, J. (2015). Experimental designs in sentence processing research: a methodological review and user's guide. *Stud. Sec. Lang. Acquisit.* 37, 1–32. doi: 10.1017/S0272263114000187
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Statist. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lago, S., Mosca, M., and Stutter Garcia, A. (2020). The role of crosslinguistic influence in multilingual processing: lexicon versus syntax. *Lang. Learn.* 71, 12412. doi: 10.1111/lang.12412
- Lenth, R. (2020). *emmeans: Estimated Marginal Means, Aka Least-Squares Means*. R package version 1.5.2-1. Available online at: <https://CRAN.R-project.org/package=emmeans> (accessed December, 2020).
- Lindqvist, C. (2009). The use of the L1 and the L2 in French L3: examining cross linguistic lexemes in multilingual learners' oral production. *Int. J. Multilingual.* 6, 281–297. doi: 10.1080/14790710902812022
- Lorenz, E., Bonnie, R. J., Feindt, K., Rahbari, S., and Siemund, P. (2019). Cross-linguistic influence in unbalanced bilingual heritage speakers on subsequent language acquisition: evidence from pronominal object placement in ditransitive clauses. *Int. J. Bilingual.* 23, 1410–1430. doi: 10.1177/1367006918791296
- MacWhinney, B. (2005). "A unified model of language acquisition," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll, and A. M. B. De Groot, (Oxford: Oxford University Press), 49–67.
- Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Foryś-Nogala, M., Kutylowska, K., et al. (2020). False friends or real friends? False cognates show advantage in word form learning. *Cognition* 206:104477. doi: 10.1016/j.cognition.2020.104477
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilingual. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marinis, T. (2010). "Using on-line processing methods in language acquisition research," in *Language Learning & Language Teaching: Vol. 27. Experimental Methods in Language Acquisition Research*, eds E. Blom, and S. Unsworth, (Amsterdam: John Benjamins Publishing Company), 139–162. doi: 10.1075/llt.27.09mar
- McLaughlin, J., Osterhout, L., and Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704. doi: 10.1038/nn1264
- Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Quantitat. Methods Psychol.* 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Poarch, G. J., and Van Hell, J. G. (2014). Cross-language activation in same-script and different-script trilinguals. *Int. J. Bilingual.* 18, 693–716. doi: 10.1177/1367006912472262
- Poort, E. D., and Rodd, J. M. (2019). Towards a distributed connectionist account of cognates and interlingual homographs: evidence from semantic relatedness tasks. *PeerJ* 7:e6725. doi: 10.7717/peerj.6725
- Prior, A., Degani, T., Awawdy, S., Yassin, R., and Korem, N. (2017). Is susceptibility to cross language interference domain specific? *Cognition* 165, 10–25. doi: 10.1016/j.cognition.2017.04.006
- Puig-Mayenco, E., González Alonso, J., and Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Sec. Lang. Res.* 36, 31–34. doi: 10.1177/0267658318809147
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ringbom, H. (1987). *The Role of the First Language in Foreign Language Learning*. Clevedon: Multilingual Matters. doi: 10.1177/026565908900500218
- Rothman, J. (2010). On the typological economy of syntactic transfer: word order and relative clause high/low attachment preference in L3 Brazilian Portuguese. *IRAL Int. Rev. Appl. Linguist. Lang. Teach.* 48, 245–273. doi: 10.1515/iral.2010.011
- Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: the typological primacy model. *Sec. Lang. Res.* 27, 107–128. doi: 10.1177/0267658310386439
- Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: timing of acquisition and proficiency considered. *Bilingual Lang. Cogn.* 18, 179–190. doi: 10.1017/S136672891300059X

- Rothman, J., and Halloran, B. (2013). Formal linguistic approaches to L3/Ln acquisition: a focus on morphosyntactic transfer in adult multilingualism. *Annu. Rev. Appl. Linguist.* 31, 51–67. doi: 10.1017/S0267190513000032
- Rothman, J., Gonzalez Alonso, J., and Puig-Mayenco, E. (2019). *Third Language Acquisition and Linguistic Transfer*. Cambridge: Cambridge University Press. doi: 10.1017/9781316014660
- Saiegh-Haddad, E., and Henkin-Roitfarb, R. (2014). “The structure of Arabic language and orthography,” in *Handbook of Arabic Literacy: Insights and Perspectives*, eds E. Saiegh-Haddad, and M. Joshi, (Dordrecht: Springer), 3–28. doi: 10.1007/978-94-017-8545-7\_1
- Sanz, C., Park, H. I., and Lado, B. (2015). A functional approach to cross-linguistic influence in ab initio L3 acquisition. *Bilingual. Lang. Cogn.* 18, 236–251. doi: 10.1017/S1366728914000285
- Schmidt, R. (2012). “Attention, awareness, and individual differences in language learning,” in *Perspectives on Individual Characteristics and Foreign Language Education*, eds W. Chan, K. Chin, S. Bhatt, and I. Walker, (Berlin: De Gruyter Mouton), 27–50. doi: 10.1515/9781614510932.27
- Sedivy, J. C. (2010). “Using eye tracking in language acquisition research,” in *Experimental Methods in Language Acquisition Research*, eds E. Blom, and S. Unsworth, (Amsterdam: John Benjamins), 115–138. doi: 10.1075/llt.27.08sed
- Singleton, D., and O’Laoire, M. (2004). “Psychotypology and the ‘L2 factor’ in cross-lexical interaction: an analysis of English and Irish influence in learner French,” in *Språk, Lärande och Utbildning i Sikte*, eds M. Bendtsen, M. Björklund, C. Fant, and L. Forsman, (Vasa: Faculty of Education, Åbo Akademi), 191–205.
- Singleton, D., and O’Laoire, M. (2005). *Cross-Lexical Interaction in Irish-English Bilinguals’ French: Further Exploration of the Psychotypology Factor*. EUROSLA-2005. Dubrovnik: EUROSLA.
- Slabakova, R. (2017). The scalpel model of third language acquisition. *Int. J. Bilingual.* 21, 651–665. doi: 10.1177/1367006916655413
- Slabakova, R., and Garcia Mayo, M. D. P. (2015). The L3 syntax-discourse interface. *Bilingual. Lang. Cogn.* 18, 208–226. doi: 10.1017/S1366728913000369
- Tajareh, M. J. (2015). An overview of contrastive analysis hypothesis. *Cumhur. Üniv. Fen Edebiy. Fakült. Bilim. Derg.* 36, 1106–1113.
- Tokowicz, N., and MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: an event-related potential investigation. *Stud. Sec. Lang. Acquisit.* 27, 173–204. doi: 10.1017/S0272263105050102
- Tokowicz, N., and Warren, T. (2010). Beginning adult L2 learners’ sensitivity to morphosyntactic violations: a self-paced reading study. *Eur. J. Cogn. Psychol.* 22, 1092–1106. doi: 10.1080/09541440903325178
- Tolentino, L. C., and Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Lang. Learn.* 64, 279–309. doi: 10.1111/lang.12048
- Ullman, M. (2005). “A cognitive neuroscience perspective on second language acquisition: the declarative/procedural model,” in *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice*, ed. C. Sanz, (Washington, DC: Georgetown University Press), 141–178.
- Valdés Kroff, J. R., Guzzardo Tamargo, R. E., and Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguist. Approac. Bilingual.* 8, 98–133. doi: 10.1075/lab.16011.val
- Voeten, C. (2019). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. R package version 1.3. Available online at: <https://CRAN.R-project.org/package=buildmer> (accessed December, 2020).
- Westergaard, M. (2019). Microvariation in multilingual situations: the importance of property-by-property acquisition. *Sec. Lang. Res.* doi: 10.1177/0267658319884116 [Epub ahead of print].
- Westergaard, M., Mitrofanova, N., Mykhaylyk, R., and Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: the linguistic proximity model. *Int. J. Bilingual.* 21, 666–682. doi: 10.1177/1367006916648859
- Williams, S., and Hammarberg, B. (2009). “Language switches in L3 production: implications for a polyglot speaking model,” in *Processes in Third Language Acquisition*, ed. B. Hammarberg, (Edinburgh: Edinburgh University Press), 28–73. doi: 10.1093/applin/19.3.295

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Abbas, Degani and Prior. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Text Complexity Modulates Cross-Linguistic Sentence Integration in L2 Reading

Sibylla Leon Guerrero<sup>1\*</sup>, Veronica Whitford<sup>2</sup>, Laura Mesite<sup>1</sup> and Gigi Luk<sup>3</sup>

<sup>1</sup>Graduate School of Education, Harvard University, Cambridge, MA, United States, <sup>2</sup>Dept. of Psychology, University of New Brunswick, Fredericton, NB, Canada, <sup>3</sup>Dept. of Educational and Counselling Psychology, McGill University, Montreal, QC, Canada

## OPEN ACCESS

### Edited by:

Anat Prior,  
University of Haifa, Israel

### Reviewed by:

Jorge R. Valdes Kroff,  
University of Florida, United States  
Brian MacWhinney,  
Carnegie Mellon University,  
United States

### \*Correspondence:

Sibylla Leon Guerrero  
sibylla\_leon\_guerrero@  
mail.harvard.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 11 January 2021

**Accepted:** 31 May 2021

**Published:** 18 June 2021

### Citation:

Leon Guerrero S, Whitford V, Mesite L  
and Luk G (2021) Text Complexity  
Modulates Cross-Linguistic Sentence  
Integration in L2 Reading.  
Front. Commun. 6:651769.  
doi: 10.3389/fcomm.2021.651769

Cross-linguistic influences (CLI) in first-language (L1) and second-language (L2) reading have been widely demonstrated in experimental paradigms with adults at the word and sentence levels. However, less is known about CLI in adolescents during naturalistic text reading. Through eye-tracking and behavioral measures, this study investigated expository reading in functionally English monolingual and Spanish (L1) - English (L2) bilingual adolescents. In particular, we examined the role of L1 (Spanish) sentence integration skills among the bilingual adolescents when L2 school texts contained challenging syntactic structures, such as complex clauses, elaborated noun phrases, and anaphoric references. Results of generalized multilevel linear regression modeling demonstrated CLI in both offline comprehension and online eye-tracking measures that were modulated by school text characteristics. We found a positive relationship (i.e., facilitation) between L1 sentence integration skills and L2 English text comprehension, especially for passages with greater clause complexity. Similar main, but not modulatory, effects of sentence integration skill were found in online eye-tracking measures. Overall, both language groups appeared to draw upon similar reading component skills to support reading fluency and comprehension when component skills were measured only in English. However, differential patterns of association across languages became evident when those skills were measured in both L1 and L2. Taken together, our findings suggest that bilingual adolescents' engagement of cross-linguistic resources in expository reading varies dynamically according to both language-specific semantic knowledge and language-general sentence integration skills, and is modulated by text features, such as syntactic complexity.

**Keywords:** bilingual (Spanish/English), cross-linguistic influence, reading comprehension, syntactic integration, second language reading, text complexity, eye-tracking

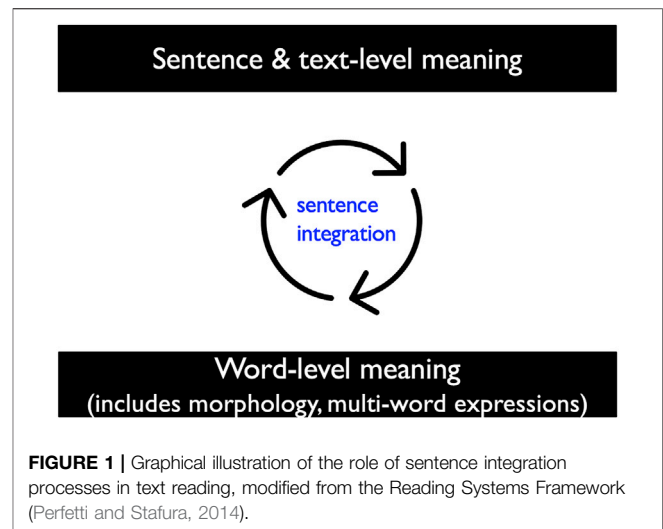
## INTRODUCTION

Reading and comprehending complex connected school texts can be challenging for both first-language (L1) (Cain and Oakhill, 2004; Lervåg et al., 2018) and second-language (L2) (Lesaux, et al., 2006; Melby-Lervåg and Lervåg, 2014) learners. Readers must draw upon linguistic knowledge and skills of different types: orthographic decoding, vocabulary, syntax, and discourse knowledge to rapidly create and update dynamic representations of meaning (Kintsch and Van Dijk, 1978; Perfetti

and Stafura, 2014). For bilinguals who are also biliterate, the reading process is further complicated by the presence of two languages, drawing upon component skills that can include more than one orthography, lexicon, and syntactic system. Cross-linguistic influence (CLI) describes the effects that bilinguals' languages may have on each other, even when only one language is the target of communication (Alonso, 2019). Cummins' (1979) early Linguistic Interdependence Hypothesis proposed a common underlying linguistic proficiency that would allow bilinguals to transfer linguistic skills across languages given sufficient L1 proficiency. This hypothesis proposes that in addition to their emerging L2 abilities, young bilinguals draw upon L1 knowledge and skills in developing L2 literacy.

Educational studies of cross-linguistic influence using behavioral measures of oral language broadly defined (e.g., Lesaux et al., 2010; Nakamoto et al., 2012; Proctor et al., 2010; Relyea and Amendum, 2020) have generally found a weak or no contribution of L1 oral language to skilled text reading and comprehension in L2, or have found that L1 contributions are mediated by L2 language skills. In contrast, extensive behavioral research has found that L1 can support L2 text reading through shared component processes, such as phonological awareness (Durgugoglu et al., 1993; Bialystok et al., 2005; Prevoo et al., 2016) and orthographic decoding (Geva and Siegel, 2000; Lindsey et al., 2003; Lesaux and Siegel, 2003; Verhoeven and van Leeuwe, 2009; Melby-Lervåg and Lervåg, 2011; Kremin et al., 2019), particularly when L1 and L2 share writing systems. Similarly, cross-linguistic transfer of L1 vocabulary knowledge, especially as children move from the stages of "learning to read" to "reading to learn," can bolster L2 reading comprehension (e.g., Proctor et al., 2005; Pasquarella et al., 2012; van den Bosch et al., 2020). However, in social contexts such as the United States, where there is often attrition of the minority home language as children's schooling progresses in the societally dominant language (i.e., English), L1 proficiency can also exert a negative influence on L2 comprehension (e.g., Swanson et al., 2008; Kieffer, 2012; Ordóñez et al., 2002). In these contexts of subtractive bilingualism (Cummins, 2000), conditions which promote L2 proficiency may also exacerbate L1 attrition, giving the impression of negative interference as higher levels of L1 are correlated with lower levels of L2 language knowledge and vice versa.

Although CLI in reading has been widely examined at the level of sound and word representations (see Genesee and Geva, 2006; Chung et al., 2019), there has been comparatively little education research on CLI involving the integration of these word representations into sentence- and text-level meanings, a higher-order process needed for skilled reading and comprehension of school texts. While rich lexical representations are important building blocks in reading, comprehension requires more than understanding words in isolation. Readers must create a context or situation model of connected concepts and rapidly integrate new words as they are read (Van Dijk and Kintsch, 1983). The Reading Systems Framework (RSF, Perfetti and Stafura, 2014) is one account of this memory-dependent integration process, which postulates that successful readers rapidly access orthographic and rich



lexical representations from written text. These rich lexical representations include morphological and other word features, such as aspect and category, that contribute to "high-quality" lexical representations. In order to comprehend a text, readers hold these representations in memory while integrating them into a holistic representation of the sentence (or sentences), called the textbase. Sentence-level integration requires the reader to both compose and decompose meaning beyond the lexical level (e.g., using word order, referential rules, and other sentence-level cues). The present study investigates this integrative processing (illustrated in **Figure 1**) as a potential locus of L1 linguistic ability that may support L2 reading efficiency and comprehension.

Here, we investigated the nature of word-to-sentence integration using naturalistic reading tasks drawn from education curricula in order to investigate the complex reading behaviors of adolescents who must learn in a developing L2. According to the RSF, linguistic knowledge contributes to the construction of meaning beyond the word level; however, the mechanism by which word-level representations are integrated at the sentence and text level is not precisely specified in the RSF. There are currently two prevalent types of models of sentence integration: memory-based integration models and predictive processing models. On the one hand, memory-based accounts of complex meaning representation have traditionally been formulated around the "bottom-up" processing of separate representations of the lexicon, morphology, and syntax that are held together in working memory to create sentence-level meaning (e.g., Cunnings, 2017). On the other hand, computational and neurobiological accounts of complex meaning representation focus on the role of "top-down" predictive or expectation-based meaning construction (e.g., Levy, 2008). Tightly controlled experimental work can provide rigorous evidence validating or disproving these types of models. However, such work also trades off experimental control and ecological validity and insight into complex behavior in rich discourse contexts (e.g., see Vanderwal et al., 2019 for a discussion of naturalistic stimuli). In order to conduct transdisciplinary educational research, the processing account in the present

naturalistic study is grounded in an ecologically relevant approach in which we adopt a top-down perspective whereby different reading component skills are revealed in behavioral and eye movement measures of reading efficiency and comprehension.

Findings are thus observational and seek to extend prior monolingual literature indicating that same-language sentence integration skills are important contributors to reading comprehension, particularly starting in late elementary and middle school (Low and Siegel, 2005; Nation and Snowling, 2000; Lesaux et al., 2006; Geva and Farnia, 2012; Proctor et al., 2012; Jeon and Yamashita, 2014; Gottardo et al., 2018; Babayigit and Shapiro, 2019; Brimo et al., 2017). In particular, little is understood about the cross-linguistic contributions of L1 sentence integration abilities when bilingual individuals are reading in their L2, a common requirement for bilingual learners with L2 English in schools with English as the medium of instruction. The current study focuses on CLI in sentence-level integration, examining how L1 (Spanish) sentence integration abilities during L2 (English) text reading in Spanish-English bilingual adolescents are associated with both online text processing efficiency (assessed by eye-tracking measures) and offline text comprehension (assessed by post-reading comprehension measures) with their functional monolingual peers who spoke and read primarily only in English.

## Educational Studies of Cross-Linguistic Influence in Text Comprehension

Although cross-linguistic influence involving the lexicon has been widely investigated, there is inconclusive evidence regarding potential CLI of syntax on L2 reading comprehension. Several behavioral studies involving children educated in L2 English-dominant environments found no relationship between L1 syntactic skills and L2 reading comprehension. In a study of adolescent Spanish-English bilinguals who were newcomers to the United States with prior schooling in Spanish, Garrison-Fletcher et al. (2019) found that general and academic reading comprehension, but not Spanish syntactic skills, were positive predictors of English text comprehension. Similarly, Swanson et al. (2008) assessed 68 third grade Spanish-English bilinguals with both Spanish and English reading component measures. Although L1 Spanish morphosyntactic knowledge positively predicted English reading comprehension, it did not do so uniquely—this relationship was explained by other English measures. With 123 children in grades 3–5, Leider et al. (2008) found a negative relationship between Spanish syntax and English sentence judgment performance but no relationship with English comprehension measured by cloze or multiple-choice tasks. Kieffer (2012) found a similar negative relationship between 295 kindergarten children's Spanish oral language skills more broadly measured and later English reading comprehension in nationally representative longitudinal data. These three studies were conducted in the United States, where the L2 (English) is the dominant language used almost exclusively in public schools. Only one study with participants in this setting produced a contrasting result. In a longitudinal

investigation of syntactic skills among 156 Spanish-English bilingual upper elementary students, Proctor et al. (2017) found that L1 syntax, measured with a sentence formulation task in the second grade, predicted English oral language and reading comprehension in the fifth grade. There is thus mixed evidence of cross-linguistic syntactic transfer in L2 English-dominant settings at different ages, with study findings variously suggesting interference, positive transfer, or no influence of L1 syntactic skills on L2 text comprehension.

In contrast, behavioral studies conducted in immersive bilingual education settings where bilingual literacy is explicitly instructed have found a consistent positive relationship between L1 syntax and L2 reading comprehension during the preschool to upper elementary years. Gabriele et al. (2009) examined a small sample of Spanish-English bilinguals in a bilingual education preschool program. Preschoolers' performance on native Spanish structures (i.e., more complex structures) predicted better performance on an English reading readiness measure. In unpublished studies, Sohail (2015) tested a cross-sectional sample of emergent Canadian English-French bilingual first graders and their third-grade peers who had completed 2 years of French immersion using a word order correction task. In the first grade, but not in the third grade, L1 English word order correction was positively associated with L2 French reading comprehension. Among bilingual Cantonese-English first through third graders studying a dual-language curriculum in Hong Kong, both cross-sectional (Siu and Ho, 2015) and longitudinal (Siu and Ho, 2020) studies found that L1 Chinese word order and morphosyntactic skills positively predicted L2 English text comprehension measured in the third grade. L2 English word order and morphosyntactic skills mediated both cross-sectional and longitudinal relationships when accounting for oral language and cognitive skills, suggesting that syntactic skills represent a common underlying linguistic proficiency contributing to reading comprehension. Similarly, for Spanish-English bilingual upper elementary students in dual immersion programs, Phillips Galloway et al. (2020) found that L2 (English) reading comprehension was positively associated with a composite measure of academic language in the L1 Spanish that included syntactic skills in addition to vocabulary and genre-related knowledge. Although each of these studies investigated different age groups (preschool to upper elementary) and different measures of syntactic skill (word order, morphosyntax, and sentence complexity), they share common findings of cross-linguistic syntactic transfer to L2 reading among children in bilingual immersion education settings. Consistent with Cummins' (1979) hypothesis that readers could draw upon a common underlying linguistic proficiency to support L2 comprehension only when they possessed adequate L1 proficiency, these findings in bilingual education settings suggest that school instruction in both L1 and L2 may support the positive influence of L1 skills on L2 reading, or at least may mitigate variability in L1 proficiency that impedes this cross-linguistic influence.

Particularly in the United States' educational context where English-only schooling has predominated in recent decades, these mixed results from studies investigating CLI may arise from

heterogeneity in either or both participant language characteristics and educational environments. For example, United States-centric studies of Spanish-English bilinguals often examine students designated in the public school system as English learners, who are predominantly United States born, but may in fact range from recent immigrants from Spanish-speaking countries with extensive Spanish-language educational experience to second generation heritage speakers with minimal Spanish proficiency (e.g. see Luk and Christodoulou, 2016 for a discussion). The present study, conducted in the United States context, seeks to understand how students for whom Spanish is the first spoken language (L1) may draw upon this L1 to support text processing and comprehension in English, their second spoken language (L2). We considered Spanish as the participant's L1 if it was the earliest language *spoken* by the participant on a daily basis at home, even for participants who may have been *exposed* to English at an early age by listening to media, community interactions, or family conversation in the context of the United States where English is the societally dominant language.

Variability found in CLI effects both within L2 English-dominant school environments and also across L2-dominant compared to immersive bilingual education settings additionally suggest that CLI is more likely to be observed in contexts where both L1 and L2 literacy are systematically taught and that skills with L1 oral vs. written syntax may have different influences on L2 written comprehension. L1 syntactic skills may thus support L2 reading comprehension; however, it is still unclear for whom and under what behavioral measures and environmental conditions this cross-linguistic support may take place.

## Eye-Tracking Studies of Cross-Linguistic Influence in Text Processing

Unlike traditional behavioral tasks, eye-tracking allows for a direct, naturalistic, and temporally sensitive measure of the cognitive processes underlying reading behaviors as they unfold (Rayner, 2009). A large body of eye-tracking literature on bilingual sentence processing has found that L1 syntactic knowledge impacts online L2 sentence reading most often by slowing text processing, for example in complex syntactic structures, such as causal connectives (van den Bosch et al., 2018), anaphora (see Godfroid, 2019 for a review), and referential clauses (see Dussias et al., 2010; Rossi et al., 2019 for a review). Eye-tracking studies of connected text reading (though not focused on CLI) have also found differences in L1 and L2 reading behavior, including longer fixations and more saccades in the L2, contributing to longer reading times. These differences can vary according to levels of self-reported L1 and L2 exposure (Whitford and Titone, 2012; Whitford and Titone, 2017) and objective L1 and L2 proficiency (Cop et al., 2015; Whitford and Joannis, 2018).

In addition to reader characteristics, effects of syntactic complexity of the text have been found in bilingual adults' adaptation to word category combinations (e.g., article-adjective-noun). Snell and Theeuwes (2020) presented a naturalistic narrative text in both Dutch and English and found that higher frequency structures facilitated Dutch-

English bilinguals' eye movement reading behavior in both languages. Differences for bilingual adults reading narrative passages in L1 vs. L2 have been found for syntactic structures such as gerunds, participial phrases, and relative clauses (De Groot, 2018). Among monolingual children, word position during expository text reading has been found to impact offline comprehension (de Leeuw et al., 2016). Furthermore, syntactic processing difficulties have been reported among monolingual children for anaphoric structures in short narrative passages (Joseph et al., 2015).

Eye-tracking studies of connected text reading have primarily focused on adults, leaving reading behavior in children relatively less understood. However, prior studies involving both monolinguals (Joseph et al., 2015; Reichle et al., 2013; Whitford and Joannis, 2018) and bilinguals (Whitford and Joannis, 2018) have reported age differences in reading performance, with children exhibiting more fixations and regressions, longer fixations, and shorter saccade amplitudes than adults (Whitford and Joannis, 2018). Although the maturation of oculomotor control appears to be largely complete by puberty, or around 12 years of age, children's eye movement reading behavior only approximates that of adults as their language proficiency and word-reading automaticity develops (Blythe and Joseph, 2011; Reichle et al., 2013). Adolescent middle schoolers in K-12 education contexts thus are likely to have adult-like oculomotor capacities as well as word-level decoding skills but are still developing in the language and higher-level complex reading skills needed for efficient text processing.

To our knowledge, studies have yet to investigate the interaction of reader characteristics, syntactic characteristics of text, and reading comprehension using both offline comprehension tasks and eye movement measures of reading in L2 adolescent readers. Kuperman et al. (2018) examined a broad range of reader and text characteristics in native English monolingual university students, and found at the passage level that syntactic complexity of passages predicted several measures of online eye movement behavior, as well as offline reading comprehension. Results for reader characteristics varied depending upon the specific eye movement measure examined, with word reading fluency and nonverbal reasoning emerging as the common predictors of online text processing. Overall in this study, there was little modulation of reader characteristics by syntactic complexity of the text when comparing less complex to more complex passages. Although prior eye-tracking studies have found that syntactic complexity of texts influences eye movement reading behavior in monolingual children (e.g., measured as sentence length in German-speaking children: Tiffin-Richards and Schroeder, 2018) and adults (e.g., English-speaking adults: Kuperman et al., 2018), none have focused on word-to-sentence integration in bilingual adolescents who are experiencing the transitional demand of learning complex academic knowledge through a developing second language.

## The Current Study

Prior CLI research, which has primarily focused on bilingual adults (and thus, lack generalizability to other populations) and

employed highly controlled sentences as stimuli (and thus, may lack ecological validity), has found that both complex syntactic features of the text and L1 proficiency influence L2 text processing and comprehension. Syntactic features prevalent in school texts, such as elaborated noun phrases, lengthy relative clauses, conceptual anaphora, and the use of distinctive connectives, have been found to pose particular challenges for written text comprehension (Scott and Balthazar, 2010; Uccelli et al., 2015a) in students generally and in L2 learners more specifically (Phillips Galloway and Uccelli, 2019). Cross-linguistically, L2 English academic skills and reading comprehension have been found to be positively associated with L1 Spanish academic vocabulary (Lubliner and Hiebert, 2011) and with academic language skills broadly measured as combined lexical, syntactic, and discourse skills (Aguilar et al., 2020; Phillips Galloway and Uccelli, 2019). However, how L2 learners process distinctive syntactic features of written school texts is not well understood, and little prior research has examined L2 syntactic processing beyond morphosyntax in written academic discourse comprehension. Concurrently, prior education research involving bilingual adolescents has found that complex syntactic text features impede L2 reading comprehension, as a function of L2 proficiency. However, there is still little work connecting these two bodies of literature and examining how syntactic characteristics of L2 texts modulate the cross-linguistic relationship of individual reader characteristics and the comprehension of naturalistic school-based texts. To address this critical gap, the current study asks: How do complex syntactic text features and individual differences in cross-linguistic sentence integration skills, as well as their interactions, affect 1) online processing and 2) offline comprehension of naturalistic school-based texts in bilingual adolescent readers?

We examined individual differences in sentence integration skills in both the L1 (Spanish) and L2 (English), as well as in word decoding and vocabulary knowledge in both languages in order to control for lexical effects. Based on prior research, we expected that better L1 and L2 sentence integration skills would be associated with higher levels of L2 text comprehension (Lesaux et al., 2006) and with more efficient online L2 reading (Cop et al., 2015; Gollan et al., 2008; Whitford and Titone, 2012; Whitford and Titone, 2015; Whitford and Titone, 2017). We included two adolescent participant groups: functionally monolingual native English speakers, and Spanish (L1) - English (L2) speakers, acquiring English in school.

Three categories of text features were included in the study: phrase complexity, clause complexity, and the degree of anaphoric reference. Based on the RSF model (Perfetti and Stafura, 2014), we expected that these features, which present challenges to sentence integration processes, such as those involved in parsing, sequencing, and combinatorial analysis/synthesis, would be a potential locus of cross-linguistic influence in which individual differences in L1 sentence integration skills might contribute to L2 text comprehension. Because syntactic complexity presents challenges to sentence integration, readers may draw upon integrative resources such as syntactic representations or processing biases developed in the

L1 to cope with these challenges. As previous research has found that complex syntactic text features are associated with reading comprehension difficulties in L2 child and adolescent readers (Uccelli et al., 2015a; Uccelli et al., 2015b; Phillips Galloway and Uccelli, 2019), we expected that phrase complexity, clause complexity, and anaphoric references would all negatively modulate the expected positive effect of cross-linguistic sentence integration skills on offline text comprehension. Based on similar findings in a self-paced L2 reading study in Spanish-English bilingual adolescents which found that words in more syntactically complex passages were read more slowly than words in simpler ones (Kim et al., 2018; Mulder et al., 2020), we expected a negative impact of these complex text features on online text processing. The study's hypotheses were that:

- 1) greater L1 proficiency in sentence integration would be associated with higher levels of L2 text reading efficiency and comprehension, and
- 2) syntactic complexity of the text would negatively modulate the expected positive association of L1 syntactic integration skills with L2 text reading efficiency and comprehension.

## MATERIALS AND METHODS

### Participants

Sixty-five typically developing adolescents, aged 11–15 years, participated in the current study. Participants resided in the Boston Metropolitan Area and attended English-instruction schools. In accordance with Institutional Review Board guidelines (Harvard University IRB16-0866), both child assent and parental consent were obtained. Participants were compensated \$50 for their time. The final sample included 59 participants; three were excluded due to eye-tracking issues and another three were excluded due to low nonverbal reasoning scores (standard scores <86.6 or > two standard deviations below the sample mean on the Kaufman Brief Intelligence Test-2; KBIT-2 Matrices; Kaufman and Kaufman, 2004), which can influence reading comprehension outcomes (Quinn and Wagner, 2018).

A parental version of the Language and Social Background Questionnaire (LSBQ; adapted from Luk and Bialystok, 2013; Anderson et al., 2018) assessed participants' demographic background, language history, and familial language use. All participants demonstrated heterogeneous language backgrounds reflective of minority language speakers found in United States classrooms. Twenty-nine participants were native L1 (English) speakers, with no L2 immersion experience and minimal proficiency in an additional language. Most were enrolled in an introductory foreign language class as part of the standard middle school curriculum, and several of these L1 English speakers had been enrolled in some form of beginning language class sporadically since an early age. Thus, they were all functionally monolingual.

We considered Spanish as the participant's L1 if Spanish was the earliest language *spoken* on a daily basis at home, even for participants who may have been *exposed* to English at an early age by listening to media, community interactions, or family

**TABLE 1** | Participant demographic characteristics.

|                                                                                                                          | L1 English (n = 29)<br>m (sd) | L1 Spanish (n = 30)<br>m (sd) | Group difference<br>( <i>p</i> of Kruskal-Wallis $\chi^2$ ) |
|--------------------------------------------------------------------------------------------------------------------------|-------------------------------|-------------------------------|-------------------------------------------------------------|
| <b>Demographics</b>                                                                                                      |                               |                               |                                                             |
| Gender (female/male)                                                                                                     | 20/9                          | 17/13                         | n/a ( <i>p</i> = 0.42)                                      |
| Maternal education (median)<br>(Categories: Elementary school; high school; 2-years (AA) degree; BA/BS; graduate degree) | Graduate degree               | Bachelor's degree             | n/a ( <i>p</i> = 0.01)                                      |
| Age (years)                                                                                                              | 12.6 (1.1)                    | 13.1 (1.0)                    | 0.54 (0.06)                                                 |
| <b>Language background</b>                                                                                               |                               |                               |                                                             |
| Age of english exposure in years (0 = "at birth")                                                                        | 0 (0)                         | 7.8 (3.7)                     | 7.8 (<0.001)                                                |
| Non-english language <sup>b</sup> spoken daily (% time at home)                                                          | 1.2 (2.7)                     | 46.9 (15.3)                   | 45.7 (<0.001)                                               |
| Non-english language <sup>b</sup> heard daily (% time at home)                                                           | 1.3 (2.7)                     | 47.3 (14.5)                   | 46.0 (<0.001)                                               |
| Non-english language <sup>b</sup> read daily (% time at home)                                                            | 1.2 (2.6)                     | 33.8 (17.5)                   | 32.6 (<0.001)                                               |

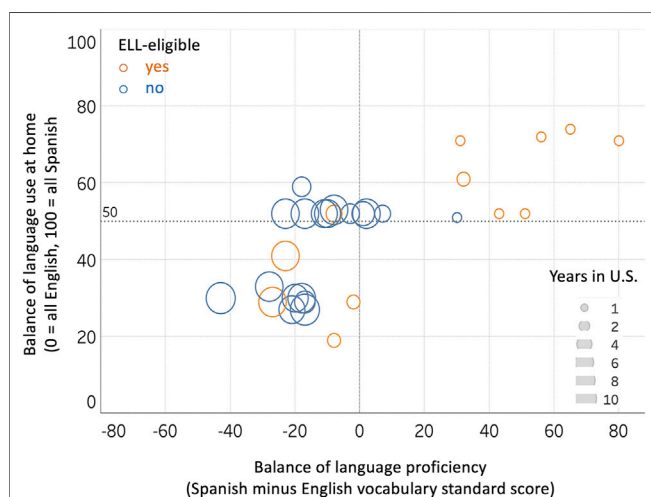
<sup>a</sup>Fisher's exact.<sup>b</sup>For L1 English speakers, this variable measured any non-English language use at home. For L1 Spanish speakers, this variable measured L1 Spanish language use at home.

**FIGURE 2** | Scatterplot showing difference scores on the x-axis, displaying the relative balance of proficiency in each language in L1 Spanish speakers (Spanish minus English standardized vocabulary scores), with zero indicating balance of proficiency, positive numbers indicating relative Spanish proficiency dominance, and negative numbers English proficiency dominance. The y-axis displays the child's language usage at home in percentage of daily use, with 50% indicating balance of language use. Marker colors indicate eligibility for English learner services, and the marker size indicates the number of years the participant has lived in the United States.

conversation. As illustrated in **Table 1**, thirty participants were identified as L1 Spanish speakers, who spoke Spanish from birth at home and used it 47% of the time, on average. Twenty-two of these participants were from Spanish-speaking countries (Mexico = 8, Central America = 5, South America = 7, Spain = 2) and eight had only lived in the United States. The age at which participants were first exposed to English on a regular basis was correspondingly heterogeneous, ranging from birth to 14 years of age ( $m = 7.8$ ,  $sd = 3.7$ ); however, all participants spoke Spanish at home and did not start to use English regularly until at least school age, or approximately 5–6 years of age.

As illustrated in **Figure 2** showing on the x-axis the difference scores between Spanish and English vocabulary measures as a

proxy for language proficiency dominance, the sample displayed a continuous range of variation on multiple and intersecting, but not fully overlapping, dimensions of language experience. For example, one participant who had lived in the United States for 10 years had markedly strong English dominance with a difference between English and Spanish vocabulary scores of over 40 points. This student could thus be considered a heritage speaker with attrition in home language but was nonetheless reported by parents to speak and read Spanish at home approximately 30% of the day and was similar in this regard to other students with English learner designations and with fewer years of residence in the United States. The multiple and at times disparate dimensions of language experience displayed in **Figure 2**—relative language proficiency, relative language usage, and relative community immersion—are reflective of the ecological variation in minority language speaker backgrounds found in United States classrooms. Also typical of United States classrooms, most of our minority language speaking sample had been initially identified as English learners upon school entry and had exited from this designation at various points in time prior to their study participation. Because initial English learner classification in United States schools is often based upon speaking a home language other than English, this designation encompasses a heterogeneous range of language proficiencies and language dominance as represented in our sample. **Figure 2** illustrates this wide distribution of students eligible for English learner services along all the other dimensions of language use, including language dominance, daily Spanish spoken, and years in the United States.

The two language groups were matched on gender (Fisher's exact  $p = 0.42$ ), age (Kruskal-Wallis  $\chi^2$   $p = 0.06$ ), nonverbal IQ (KBIT Matrices, Kaufman and Kaufman, 2004, Kruskal-Wallis  $\chi^2$   $p = 0.07$ ), verbal working memory (Digits Backward, Reynolds and Voress, 2007, Kruskal-Wallis  $\chi^2$   $p = 0.05$ ), and rapid naming (Letter, Number, 2-set, and 3-set subtests; Wolf and Denckla, 2005; Kruskal-Wallis  $\chi^2$   $p = 0.16$ ). These tasks were administered in the participant's preferred language. The two language groups were also matched on timed English word reading (Sight Word Efficiency, Torgesen et al., 2011, Kruskal-Wallis  $\chi^2$   $p = 0.96$ ) and untimed English word reading (Word Identification, Woodcock-Muñoz Language Survey with Normative Update, WMLS-R-NU,

**TABLE 2 |** Reading and cognitive characteristics.

|                                                                 | L1 English (n = 29)<br>m (sd) | L1 Spanish (n = 30)<br>m (sd) | Group difference<br>(p of Kruskal-Wallis $\chi^2$ ) |
|-----------------------------------------------------------------|-------------------------------|-------------------------------|-----------------------------------------------------|
| <b>Administered in preferred language (Spanish or English):</b> |                               |                               |                                                     |
| KBIT-2 nonverbal reasoning <sup>a</sup>                         | 114.6 (10.7)                  | 109.7 (10.1)                  | 5.0 (0.07)                                          |
| TOMAL digits backward <sup>b</sup>                              | 10.9 (3.4)                    | 9.5 (2.4)                     | 1.5 (0.05)                                          |
| RAN/RAS (subtest average) <sup>a</sup>                          | 107.8 (10.1)                  | 111.1 (10.2)                  | -3.3 (0.16)                                         |
| TOWRE-2 sight word fluency <sup>a</sup>                         | 108.6 (14.9)                  | 107.6 (15.6)                  | 1.0 (0.96)                                          |
| <b>Administered in English:</b>                                 |                               |                               |                                                     |
| WMLS-R English picture vocabulary <sup>a</sup>                  | 113.4 (10.2)                  | 84.6 (23.4)                   | 28.8 (<0.001)                                       |
| WMLS-R English letter-word ID <sup>a</sup>                      | 114.5 (12.1)                  | 111.9 (11.1)                  | 2.7 (0.46)                                          |
| WMLS-R English cloze <sup>a</sup>                               | 105.5 (14.6)                  | 87.1 (22.3)                   | 18.4 (0.004)                                        |
| <b>Administered in Spanish:</b>                                 |                               |                               |                                                     |
| WMLS-R Spanish vocabulary <sup>a</sup>                          | —                             | 87.1 (12.1)                   | —                                                   |
| WMLS-R Spanish letter-word ID <sup>a</sup>                      | —                             | 127.9 (17.6)                  | —                                                   |
| WMLS-R Spanish cloze <sup>a</sup>                               | —                             | 88.6 (14.1)                   | —                                                   |

<sup>a</sup>Standardized test with  $M = 100$  and  $SD = 15$ .<sup>b</sup>Standardized test with  $M = 10$  and  $SD = 3$ .

Woodcock et al., 2005; Kruskal-Wallis  $\chi^2 p = 0.46$ ). Median maternal education was higher in the L1 English group (graduate degree) than the L1 Spanish group (bachelor's degree, Fisher's exact  $p = 0.01$ ). L1 English speakers also performed higher on average than L1 Spanish speakers on measures of English vocabulary knowledge (WMLS-R-NU Picture Vocabulary, Woodcock et al., 2005; Kruskal-Wallis  $\chi^2 p = 0.46$ ). Measures of Spanish proficiency (*Vocabulario sobre dibujos*, *Identificación de letras y palabras*, and *Comprensión de textos*, Woodcock et al., 2005) were only administered to L1 Spanish speakers due to L1 English speakers' reported lack of Spanish proficiency and language experience. L1 Spanish speakers' English and Spanish skills were balanced on average, with no significant difference between English and Spanish vocabulary (Wilcoxon signed-rank  $p = 0.76$ ) or cloze ( $p = 0.77$ ) scores. Word identification skills for this group were higher on average in Spanish than English ( $p < 0.001$ ) although English scores on this subtest remained significantly above the population mean of 100 ( $p < 0.001$ ). See **Table 2** for participants' cognitive and language proficiency characteristics and **Supplementary Table SA**, for psychometric information on the standardized assessments.

## Measures

### Cloze Integration Task

Cloze tasks are commonly used as measures of word predictability, as well as individual differences in vocabulary, prediction, and syntactic skills. These tasks were originally developed as an indicator of reading comprehension (Taylor, 1953) and are frequently used as such in both L1 (see Collins et al., 2018 for a meta-analysis) and L2 (Tremblay, 2011; Trace, 2020) education research and practice. However, there is ongoing debate regarding the precise skills that cloze tasks assess, with studies of concurrent validity suggesting that these measures do not align well with other forms of reading comprehension assessment, such as post-reading questioning, whether in multiple-choice or open-ended format (Cutting and Scarborough, 2006; Francis et al., 2006; Keenan et al., 2008; Keenan and Meenan, 2014). There is broad consensus in the

research literature that the manner in which the task is constructed strongly influences the skills it taps for both L1 (Gellert and Elbro, 2013) and L2 speakers (Alderson, 1979; Kleijn et al., 2019). At least for L1 speakers, multi-sentence thematic contexts tap into global, discourse knowledge (Clark and Kamhi, 2014), while cloze assessments such as the WMLS-R and other Woodcock passage comprehension formats with single-sentence stimuli and more weakly constraining thematic contexts draw more strongly into lexical knowledge (Leider et al., 2013), word familiarity (Cutting and Scarborough, 2006; Francis et al., 2006), and syntactic knowledge (Cutting and Scarborough, 2006; Keenan et al., 2008; Deacon and Kieffer, 2018).

We employed a cloze task to assess participants' sentence integration abilities in each language. We administered the English WMLS-R-NU Passage Comprehension to all participants and the Spanish *Comprensión de textos* to L1 Spanish speakers only (Woodcock et al., 2005). The Spanish subtest is a parallel and equated form of the WMLS-R English Passage Comprehension subtest (Woodcock et al., 2005). This cloze task asks the participant to read one or two sentences and orally supply a single missing word. The task starts with stimuli consisting of a single, very brief sentence, expanding to two sentences and/or a longer sentence stimulus as the difficulty of the task increases. Thus, the WMLS-R-NU cloze tasks primarily assess word-to-text integration at the sentence level, particularly for developmental levels of language proficiency. All regression analyses of the cloze tasks also included English and Spanish vocabulary and word identification scores to control for the role of lexical knowledge and word familiarity in sentence comprehension (Francis et al., 2006). Age-scaled standard scores were used in analysis.

### Experimental Stimuli

Stimuli design paralleled that of Whitford and Joannis (2018) and consisted of four expository paragraphs (~100 words each), taken from the Basic Reading Inventory (BRI), 10th ed, a grade-leveled reading inventory commonly used in schools to evaluate reading fluency and comprehension (Johns, 2008). One

**TABLE 3 |** Word characteristics for passages at BRI text grade levels 5 through 8.

|                                                           | Text level 5 | Text level 6 | Text level 7 | Text level 8 |
|-----------------------------------------------------------|--------------|--------------|--------------|--------------|
| Word count                                                | 99           | 100          | 100          | 100          |
| Spanish cognates                                          | 28           | 25           | 26           | 26           |
| Lexile level                                              | 680L         | 800L         | 1010L        | 1020L        |
| SUBTLEX-US word frequency (log), mean, content words only | 3.02         | 3.33         | 3.66         | 3.51         |
| SUBTLEX-US word frequency (log), mean, all words          | 4.20         | 4.35         | 4.41         | 4.42         |
| Mean word length (letters)                                | 4.15         | 4.25         | 4.82         | 4.78         |
| Mean word length (syllables)                              | 1.31         | 1.32         | 1.47         | 1.47         |
| Mean word AoA rating                                      | 5.01         | 4.71         | 5.29         | 5.19         |
| Mean word predictability                                  | 0.31         | 0.32         | 0.24         | 0.24         |

paragraph was selected at each BRI level of difficulty corresponding to fifth to eighth grades. Paragraph topics focused on the natural world (ocean, plants, sunflowers, and the environment) and social sciences (indigenous people of the Americas). The BRI itself does not provide validity and reliability statistics for the 10th edition; however, independent analysis of BRI 10th and 11th ed. fluency measures with a younger sample (third through fifth graders) provide some evidence of high test-retest reliability (Pearson  $r = 0.89$  to  $0.93$ ), alternate forms reliability (Pearson  $r = 0.84$  to  $0.96$ ), and concurrent validity with the DIBELS (Pearson  $r = 0.85$  to  $0.97$ , Bieber et al., 2015), a screening assessment also widely used in schools (Good and Kaminsky, 2002). The six factual open-ended comprehension questions were drawn from a set of ten provided in the BRI for each passage and orally administered. Participants responded to questions orally and received dichotomous scores (correct = 1; incorrect = 0) for each comprehension question following the BRI scoring procedure (minimum score = 0 and maximum score = 6 per paragraph).

For each paragraph, we obtained lexical features known to influence reading behavior (Clifton et al., 2007): word age of acquisition (AoA), word frequency, word predictability, orthographic and phonological length, and the number of Spanish cognates (Table 3). Word AoA values were derived from Brysbaert and Biemiller (2017) test-based word AoA ratings. Word frequencies were obtained from the SUBTLEX-US corpus (Brysbaert and New, 2009). Orthographic and phonological length, as well as Levenshtein distances for cognates, were calculated using the R (R Core Team, 2013) package *stringdist* (van der Loo, 2014). Word predictabilities were obtained through a computerized cumulative cloze task (following Whitford and Titone, 2012; Whitford and Titone, 2014; Whitford and Titone, 2017), where a separate sample of adult L1 English speakers ( $n = 30$ ), guessed the words of each paragraph on a word-by-word basis. Accuracy scores were averaged across participants to create a word-level probability of cloze prediction. These word-level features were next averaged over paragraphs to produce mean characteristics for these linguistic units. As seen in Table 3, paragraphs had approximately the same number of words and Spanish cognates. Texts differed only in the average word frequency of their content words ( $p = 0.02$ ) with the lowest log frequency mean in text level 5 and the highest in text level 7. Texts did not differ in

average word frequency, predictability, AoA rating, length, proportion of content to function words, or syllable length (all Kruskal-Wallis  $p > 0.05$ ).

### Paragraph Syntactic Complexity

Each paragraph was evaluated using three classes of syntax measures: 1) classic measures that employ length, or number of words, as an indicator of syntactic complexity; 2) word-level syntax measures; and 3) syntax measures beyond the word level (e.g., examining phrases, clauses, sentences, etc.). Most sentences were composed of a single utterance called a T-unit, itself containing only a single clause. Classic syntactic complexity measures were first generated using the L2 Syntactic Complexity Analyzer (L2SCA; Lu, 2010). L2SCA output was reviewed manually. The greatest numeric variability across paragraphs is captured by sentence length (in number of words), verb phrases, coordinate phrases, and complex nominals. See **Supplementary Table SB**, for more on L2SCA measures. Next, word-level syntax measures were generated using the Tool for the Automatic Analysis of Cohesion 2.0 (TAACO; Crossley et al., 2019), which employs the Stanford Natural Language Parser (NLP). The word-level parses were inspected for errors manually and corrected where needed. TAACO analysis produces part-of-speech (POS) information as well as type-token measures for key function words at the lemma level. Semantic similarity is represented by Word2Vec (Mikolov et al., 2013), a neural network-derived representation of semantic distance among words and phrases. **Supplementary Table SC**, provides further information on the TAACO lexico-syntactic complexity measures. Finally, syntactic complexity measures beyond the word level were generated using the Tool for the Automatic Analysis of Syntactic Complexity (TAASC; Kyle and Crossley, 2018), which computes complexity indicators using Stanford NLP parses. Unlike the TAACO measures, TAASC provides syntactic complexity indicators that reflect relationships among words in a sentence and across sentences in a paragraph. TAASC indicators of syntactic sophistication (average lemma log frequency and average verb-argument construction log frequency) measure the overlap with lemmas and constructions found in the academic sub-corpus of the Corpus of Contemporary American English (COCA; Davies, 2009, 2010 as cited in Kyle and Crossley, 2018). See **Supplementary Table SD**, for more on TAASC phrase-level syntactic complexity measures.

Phrase elaboration, clause complexity, and anaphoric reference were quantified by conducting a nonlinear principal components analysis (PCA) for each characteristic. The nonlinear PCA for phrase complexity incorporated phrase measures from the parsing tools described above: adjective modifiers, nominal dependents, direct and prepositional object dependents, and prepositional phrases as nominal modifiers. The nonlinear PCA for clause complexity included clause measures: relative clause modifiers, dependent clauses, clausal direct objects, and clausal conjunctions. For anaphoric reference, the variables entered into PCA analysis were pronoun density, pronoun-noun ratio, and demonstratives. The first principal component from each of these PCA analyses was used as a corresponding text characteristic predictor in subsequent linear regression models. For phrase complexity, the first principal component accounted for 91% of total variance, for clause complexity 93%, and for anaphoric relations, 67%. See **Supplementary Table SE**, for details.

## Design and Procedure

After consenting, participants completed the eye-tracking task, where they silently read one practice paragraph and four experimental paragraphs at their own pace. After reading each paragraph, they answered six comprehension questions without being able to refer back to the passage. The order of the experimental paragraphs was counterbalanced across the two language groups in a Latin Square design. Lastly, participants completed the behavioral tasks, which were presented in a fixed order: KBIT-2, RAN/RAS, TOWRE-2, English WMLS-R, and then TOMAL-2 for L1 English speakers and KBIT-2, RAN/RAS, Spanish WMLS-R, TOMAL-2, English WMLS-R, and TOWRE-2 for L1 Spanish speakers. Total participation duration was about 3 hours.

## Language of Testing

Instructions for the nonverbal reasoning task (KBIT-2) as well as the full measures of lexical access/naming speed (RAN/RAS) and working memory (TOMAL-2) were administered in the participant's preferred language. Thirteen out of thirty L1 Spanish speakers chose to complete the RAN/RAS in English, and ten out of thirty did so for the TOMAL-2. English language measures (WMLS-R English) were administered in English only, and Spanish language measures (WMLS-R Spanish) were administered in Spanish only.

## Eye-Tracking Procedure

Participants binocularly viewed single paragraphs displayed in yellow text (14 pt. Courier New font, double-spaced) on a black background using Experiment Builder (SR Research Ltd., Ontario, Canada). Each paragraph was presented in its entirety on a 1,024 px × 786 px 21-in. screen positioned 70 cm from the participants, who maintained a fixed head position with the aid of a chin-rest. An EyeLink 1,000 desk-mounted eye-tracker (SR Research Ltd., Ottawa, Canada) collected right-eye monocular data at a 1000 Hz sampling rate. Calibration was performed before the start of each paragraph using a 9-point grid and repeated as necessary until the average fixation error was less than .5° of visual angle.

## Eye-Tracking Data Preprocessing

Trial data were inspected and corrected for vertical drift. Fixations shorter than 80 ms and those outside of word-level interest areas were deleted from the base data. No upper bound was applied to fixation durations; the maximum fixation duration was 3,330 ms (the only duration above 3,000 ms). The next seven largest fixations fell between 2000 and 2,550 ms with all remaining durations shorter than 2000 ms.

Fifteen eye-movement measures were examined based on prior literature (Cop et al., 2015; Rayner, 2012; Whitford and Joannis, 2018; Whitford and Titone, 2012; Whitford and Titone, 2017). Six of these were early stage, local (word-level) measures, which captured unconscious processing of the text during the first reading of the paragraph, also called the first pass or first run: first fixation duration and gaze duration, first pass mean saccade amplitude, first pass regressions out, and first pass word skipping. Nine late-stage eye-movement measures captured conscious integration of information and included all passes through the text: five at a local (word) level (mean fixation duration, total reading time, fixation count, regressions out, and mean saccade amplitude), and the remaining four at a global (trial) passage level (trial fixation count, saccade count, run count, and total trial time). Of these nine late measures, five (total fixation duration, total reading time, regressions out and run count and mean saccade amplitude) provided insight into late processing in specific, local areas of interest while four (total trial fixation count, saccade count, words skipped and trial duration) were indicators of global, or paragraph-level, text processing (see **Table 4** for calculation of these measures).

As illustrated by these related measures, eye-tracking output produces high-dimensional data with resulting analytic challenges. On the one hand, the multiple eye measures provide different insights into the timing of cognitive processing. Particularly for syntactic processing, the cognitive processing may be observed in some measures and not others (Clifton et al., 2007; Rayner, 1998). On the other hand, analysis with multiple, correlated outcomes or predictors augment the likelihood of type I error in eye-tracking analysis (von der Malsburg and Angele, 2017). Data-driven dimensional reduction techniques provide one way to navigate the problems of dimensionality and multiple comparisons (Kuperman et al., 2018). The current study thus employed ordinal principal components analysis (*princals* in the *Gifi* R package, Mair et al., 2017) to extract shared variance from eye-tracking measures; the first principal component from the analysis then served as the single outcome variable representing *reading efficiency* in subsequent regression analysis. Lower values indicate more efficient reading performance.

## Analytic Methods

### Descriptive Statistics

Inspection of the raw data revealed that in most cases, the distribution of behavioral and eye-tracking data violated assumptions of normality and variance homogeneity, and in regression models, of sphericity and homoscedasticity. This analysis thus utilized non-parametric tests implemented in R (R Core Team, 2018) to compute basic descriptive correlations and first-level group comparisons. Specifically, the analysis

**TABLE 4 |** Definition of eye-tracking measures.

| Measure                      | Description                                                                    |
|------------------------------|--------------------------------------------------------------------------------|
| <b>Early measures</b>        |                                                                                |
| First fixation duration      | Duration of the 1st fixation on a word                                         |
| Gaze duration                | Summed duration of all 1st pass fixations on a word                            |
| First pass saccade amplitude | Amplitude of the 1st saccade entering a word in degrees of visual angle        |
| First pass fixation count    | Number of fixations on a word during the first pass through the text           |
| First pass regression count  | Number of regressions out of a word during the first pass through the text     |
| Word skipping                | Whether a word was skipped during the first pass (yes/no)                      |
| <b>Late measures</b>         |                                                                                |
| Mean fixation duration       | Mean duration of all fixations on a word                                       |
| Total reading time           | Summed duration of all fixations on a word                                     |
| Fixation count               | Number of fixations on a word during the trial                                 |
| Regression out               | Number of regressions out of a word during the trial                           |
| Mean saccade amplitude       | Mean amplitude of all saccades entering a word (in degrees of visual angle)    |
| Trial fixation count         | Total number of fixations during the trial                                     |
| Trial saccade count          | Total number of saccades in the trial (paragraph)                              |
| Run count                    | Total number of passes or runs through the text in the trial                   |
| Total trial time             | Total trial time, from start to when participant presses a key to stop reading |

employed Wilcoxon (Mann Whitney) signed rank tests (R package *coin*; Zeileis et al., 2008) and BCa (bias corrected and adjusted) bootstrapped 95% confidence intervals to test differences in sample means (R package *boot*, Canty, 2002); Kendall's test of association (tau) to examine pairwise correlations among numeric variables of interest (in *base R*); one-way Kruskal-Wallis ANOVAs to examine associations between numeric and categorical variables of interest (in *base R*); and repeated measures, robust ANOVA (R package *WRS2*; Mair and Wilcox, 2016).

### Linear Regression Modeling

In order to construct a regression taxonomy for online eye-tracking and offline comprehension outcomes separately, the best-fitting distribution of the outcome variable was first determined through visual inspection, substantive alignment and likelihood ratio tests in R package family *gamlss* (Rigby and Stasinopoulos, 2005). The model taxonomy for analysis comprised a set of multilevel linear regression models using this best-fit outcome distribution with crossed random intercepts for subjects at the paragraph level through R package *lme4* (Bates et al., 2015) and R package family *gamlss* (Stasinopoulos et al., 2017). To examine autocorrelation effects in the eye-tracking data (Baayen et al., 2017), the linear baseline model in each taxonomy was also fitted with cubic polynomial and p-spline smoothers for the behavioral variables of interest using *gamlss*. In each case, BIC model evaluation indicated that the linear model provided the best fit to the data, hence all taxonomies represent linear mixed effects models.

For the taxonomy examining paragraph comprehension as outcome, model predictors were selected in three steps: first, the a priori selection of age, maternal education, and nonverbal reasoning were entered as control variables based on established relationship between these variables and reading outcomes (Burchinal et al., 2002; Hoff, 2013; Peterson et al., 2017; Auerbach et al., 2019; Sorenson Duncan and Paradis, 2020). Second, the full predictor dataset was reduced through

bidirectional stepwise regression minimizing generalized AIC (GAIC). Finally, variables identified in step two were corroborated using ridge regression with 10-fold cross-validation. Optimal model lambda was identified as the penalization factor yielding the lowest mean-squared error out of a range from 0.1 to 50. Ridge regression was considered to corroborate stepwise regression results if variables with ridge coefficients (i.e., effect sizes) greater than 0.10 were the same as those retained in the stepwise regression.

The model taxonomy examining reading efficiency as outcome included: 1) a baseline model in the full L1 English and L1 Spanish sample with word length, frequency, and predictability as control variables and scaled English behavioral predictors; 2) scaled text characteristic predictors were then added to the full sample model and replicated in 3) the sample of L1 Spanish speakers only. The final eye-tracking model for the L1 Spanish sample included scaled Spanish language behavioral assessments and statistically significant interactions between the Spanish language measures and text syntactic characteristics. Residual plots (residuals vs. fitted values, quantile, and residuals vs. leverage) for all models were examined to ensure that model assumptions were not violated.

## RESULTS

### Reading Comprehension

Our first hypotheses concerned cross-linguistic influence of component reading processes on offline comprehension of naturalistic English school texts of varying syntactic complexity. We first identified patterns in children's paragraph comprehension and the association of these patterns with specific syntactic characteristics of the paragraphs using multidimensional scaling (Borg et al., 2012). Next, behavioral measures that characterized paragraph comprehension outcomes were jointly plotted with these syntactic characteristics. Finally,

**TABLE 5 |** Paragraph comprehension accuracy.

|                | Combined (n = 59)<br>m (sd) | L1 English (n = 29)<br>m (sd) | L1 Spanish (n = 30)<br>m (sd) | Group mean difference<br>(p of Kruskal-Wallis $\chi^2$ ) |
|----------------|-----------------------------|-------------------------------|-------------------------------|----------------------------------------------------------|
| All paragraphs | 0.68 (0.19)                 | 0.72 (0.17)                   | 0.63 (0.21)                   | 0.09 (0.11)                                              |
| Level 5        | 0.76 (0.21)                 | 0.82 (0.18)                   | 0.71 (0.23)                   | 0.11 (0.03)*                                             |
| Level 6        | 0.78 (0.29)                 | 0.83 (0.21)                   | 0.74 (0.35)                   | 0.09 (0.69)                                              |
| Level 7        | 0.62 (0.26)                 | 0.67 (0.26)                   | 0.57 (0.25)                   | 0.10 (0.09)                                              |
| Level 8        | 0.54 (0.26)                 | 0.56 (0.30)                   | 0.52 (0.22)                   | 0.04 (0.60)                                              |

\* $p < .05$ .**TABLE 6 |** Results of fitting a taxonomy of multi-level gaussian regression models for English paragraph comprehension scores as a function of control variables age, nonverbal reasoning and maternal education, and predictor variables English word fluency, English vocabulary, English cloze, Spanish vocabulary, and Spanish cloze accuracy, n = 59 middle-schoolers, displaying standardized coefficients).

|                                                                | Model 1.1<br>$\beta$ (se) | Model 1.2<br>$\beta$ (se) | Model 2.1<br>$\beta$ (se) | Model 2.2<br>$\beta$ (se) | Model 2.3<br>$\beta$ (se) |
|----------------------------------------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Intercept                                                      | 0.67 (0.03)***            | 0.64 (0.02)***            | 0.73 (0.02)***            | 0.94 (0.07)***            | 0.92 (0.06)***            |
| <b>Control variables</b>                                       |                           |                           |                           |                           |                           |
| Age                                                            | 0.02 (0.01)               | —                         | —                         | —                         | —                         |
| Nonverbal reasoning                                            | 0.11 (0.01)***            | 0.03 (0.01)*              | 0.00 (0.02)               | 0.02 (0.02)               | —                         |
| Maternal education (0 = pre-BA/BS; 1 = BA/BS<br>2 = MA/MS/PhD) | 0.03 (0.04)               | —                         | —                         | —                         | —                         |
| Group (1 = Spanish)                                            | -0.00 (0.04)              | —                         | —                         | —                         | —                         |
| <b>Text-level predictors</b>                                   |                           |                           |                           |                           |                           |
| Phrase elaboration                                             | —                         | 0.00 (0.03)               | 0.02 (0.04)               | —                         | —                         |
| Clause complexity                                              | —                         | -0.10 (0.03)***           | -0.07 (0.04)              | -0.09 (0.02)***           | -0.27 (0.05)***           |
| Anaphoric reference                                            | —                         | -0.00 (0.01)              | -0.01 (0.02)              | —                         | —                         |
| <b>Participant-level predictors</b>                            |                           |                           |                           |                           |                           |
| English word fluency                                           | —                         | 0.01 (0.01)               | -0.04 (0.02)              | -0.01 (0.03)              | —                         |
| English vocabulary                                             | —                         | 0.08 (0.03)**             | 0.11 (0.03)***            | 0.10 (0.04)**             | 0.13 (0.02)***            |
| English cloze                                                  | —                         | 0.06 (0.02)*              | 0.07 (0.03)*              | 0.03 (0.04)               | —                         |
| Spanish word reading                                           | —                         | —                         | —                         | -0.02 (0.08)              | —                         |
| Spanish vocabulary                                             | —                         | —                         | —                         | -0.47 (0.12)***           | -0.51 (0.10)***           |
| Spanish cloze                                                  | —                         | —                         | —                         | 0.26 (0.10)**             | 0.30 (0.08)***            |
| Clause complexity X Spanish cloze                              | —                         | —                         | —                         | —                         | 0.19 (0.05)***            |
| $N_{ID}$                                                       | 59                        | 59                        | 30                        | 30                        | 30                        |
| Variance <sub>ID</sub>                                         | 0.01                      | 0.01                      | 0.01                      | 0.00                      | 0.00                      |
| Observations                                                   | 235                       | 235                       | 119                       | 119                       | 119                       |
| AIC                                                            | 1.3                       | -68.4                     | -35.8                     | -38.2                     | -58.4                     |
| BIC                                                            | 132.6                     | 69.8                      | 34.1                      | 18.7                      | -0.03                     |

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

regression taxonomies using behavioral predictors modeled paragraph comprehension outcomes.

## Descriptive Statistics

On average for each paragraph, participants answered four out of the six open-ended questions correctly (mean accuracy = 0.68, sd = 0.19). As **Table 5** illustrates, scores in the full sample were numerically lower for level 7 and 8 paragraphs than for levels 5 and 6, but scores on level 5 and 6 (post-hoc Hochberg family wise error correction,  $p = 0.58$ ) and on level 7 and 8 (Hochberg  $p = 0.06$ ) paragraphs did not differ significantly from each other. When examined pairwise by paragraph without multiple comparisons correction, accuracy means differed across language groups only for the level 5 paragraph (Kruskal-Wallis  $\chi^2 p = 0.03$ ). Variability in scores differed

across paragraphs (Mauchly sphericity test  $p = <0.001$ ), but not across groups (Mauchly's  $p = 0.09$ ), and there was no group by paragraph interaction (Mauchly's  $p = 0.77$ ).

## Linear Regression Modeling

Because multidimensional scaling is primarily an exploratory method that quantifies and visualizes dissimilarities among object scores, which in this study represented participant's accuracy in responding to paragraph questions, we next fit a taxonomy of regression models to the data in order to examine associations among behavioral predictors and paragraph comprehension in a linear regression framework. The first three models in the regression taxonomy focused solely on English language. The final model additionally incorporated behavioral measures of Spanish language skill in asking

whether cross-linguistic vocabulary knowledge and cloze abilities explained variance in paragraph comprehension over and above English measures.

As L1 Spanish and English speakers differed on average in levels of maternal education, the first model in taxonomy one included the control variables of age, nonverbal reasoning (KBIT-2), and maternal education. As displayed in **Table 6**, Model 1.1, no coefficients for age nor any level of maternal education reached significance (all  $p > 0.05$ ), while nonverbal reasoning was positively associated with paragraph comprehension ( $\beta = 0.11$ ,  $p < 0.001$ ).

Model 1.2 next determined whether paragraph syntactic characteristics and behavioral measures predicted English paragraph comprehension when controlling for nonverbal reasoning, the only significant control variable from Model 1.1. Reading component predictors were word reading fluency, vocabulary and cloze performance, all in English. The three syntactic characteristics of paragraphs in **Table 5** taxonomy, phrase and clause complexity, and anaphoric reference were quantified by conducting a nonlinear principal components analysis (PCA) for each characteristic.

As displayed in Model 1.2, L1 Spanish speakers overall ( $\beta = 0.07$ ,  $p = 0.02$ ) scored higher on average on paragraph responses when controlling for paragraph syntactic characteristics, nonverbal reasoning, English word fluency, English vocabulary, and English cloze. English vocabulary ( $\beta = 0.08$ ,  $p = 0.003$ ) and cloze ( $\beta = 0.06$ ,  $p = 0.01$ ) were positively associated with paragraph comprehension. Clause complexity was negatively related to paragraph response accuracy such that paragraphs with higher clause complexity were associated on average with lower accuracy ( $\beta = -0.10$ ,  $p < 0.001$ ). Phrase complexity ( $\beta = 0.001$ ,  $p = 0.96$ ) and anaphoric relations ( $\beta = -0.003$ ,  $p = 0.82$ ) were not significantly associated with paragraph comprehension.

Because prior research provides conflicting findings on the role of word reading performance in L2 readers, Model 2.1 next examined the Model 1.2 predictors in a reduced sample of only L1 Spanish speakers. Model 2.1 demonstrates that across the subsample and when considering English language performance and cognitive and demographic measures, results mirror those of the full sample, with English vocabulary ( $\beta = 0.11$ ,  $p = 0.001$ ) and cloze ( $\beta = 0.07$ ,  $p = 0.02$ ) as the only significant behavioral predictors of accurate paragraph comprehension when controlling for nonverbal reasoning and English word fluency abilities.

Model 2.2 then determined whether Spanish vocabulary and cloze accuracy additionally contributed to explaining variance in English paragraph comprehension for L1 Spanish speakers. When these predictors were entered into a new model that excluded the non-significant syntactic characteristics in Model 2.1, English vocabulary positively predicted paragraph comprehension ( $\beta = 0.10$ ,  $p = 0.007$ ) just as in the prior model, while Spanish vocabulary was negatively associated with comprehension ( $\beta = -0.47$ ,  $p < 0.001$ ), such that higher levels of Spanish vocabulary knowledge predicted lower accuracy on the comprehension questions. In contrast to Spanish vocabulary, however, Spanish cloze accuracy in Model 2.2

indicated a positive relationship, such that stronger Spanish cloze abilities predicted greater paragraph comprehension in L2 English ( $\beta = 0.26$ ,  $p = 0.009$ ) when controlling for English and Spanish vocabulary levels. Furthermore, as seen in the final model of the taxonomy, Model 2.3, which removes non-significant predictors in the interest of parsimony, there was a significant interaction, illustrated in **Figure 3**, between Spanish cloze scores and clause complexity ( $\beta = 0.19$ ,  $p < 0.001$ ), such that higher Spanish cloze scores attenuated the negative association of paragraph clause complexity with paragraph comprehension.

## Eye Movement Reading Behavior

Using syntactic elements identified based on the text analysis, we next asked what kind of relationship between eye movement measures, syntax, and L1 and L2 reading skills is observed in adolescents when reading naturalistic school texts in English. This second set of analyses identified patterns in adolescents' eye-tracking measures and the association of these patterns with specific syntactic characteristics of the stimulus paragraphs using multidimensional scaling (Borg et al., 2012). Next, behavioral measures that characterized paragraph comprehension outcomes were jointly plotted with these syntactic characteristics. Finally, regression taxonomies using syntactic and behavioral predictors were used to model eye-tracking outcomes.

## Descriptive Statistics

Overall, first fixation and mean fixation durations at the word level were comparable for both language groups (Kruskal-Wallis  $p = 0.87$ ). In addition, they did not differ in either first pass ( $p = 0.84$ ) or total ( $p = 0.96$ ) regressions at the word level. However, as **Table 7** demonstrates, and consistent with prior findings on L2 reading (Whitford and Joannis, 2018; Cop et al., 2015), L1 Spanish speakers engaged in significantly more fixations on average at the word ( $p = 0.001$ ) and trial ( $p = 0.002$ ) levels, contributing to a concomitantly longer average trial time ( $p = 0.003$ ). L1 Spanish speakers also skipped fewer words, on average, in both first pass ( $p = 0.008$ ) and total reading ( $p = 0.001$ ), and their saccades were correspondingly shorter in both first pass ( $p = 0.002$ ) and total reading ( $p = 0.002$ ) than those of their L1 English counterparts. The group differences displayed on paragraphs overall also held true for paragraphs when examined separately (all uncorrected Kruskal-Wallis  $p < 0.05$ ).

## Regression Modeling

Given the substantively interrelated nature of these measures and their high correlation, as well as loadings on a single PCA dimension, the eye-tracking measurement regression taxonomies used the first principal components dimension, accounting for 53.4% of variance in the 15 eye-tracking measures described above as the outcome variable. All eye measurement variables loaded on this first dimension such that variables typically positively associated with faster and more proficient reading (e.g., word-skipping, saccade amplitude) loaded with opposite sign to variables typically negatively associated with proficient reading (e.g., gaze durations, regressions, number of fixations), which loaded negatively. Secondary analysis using single eye

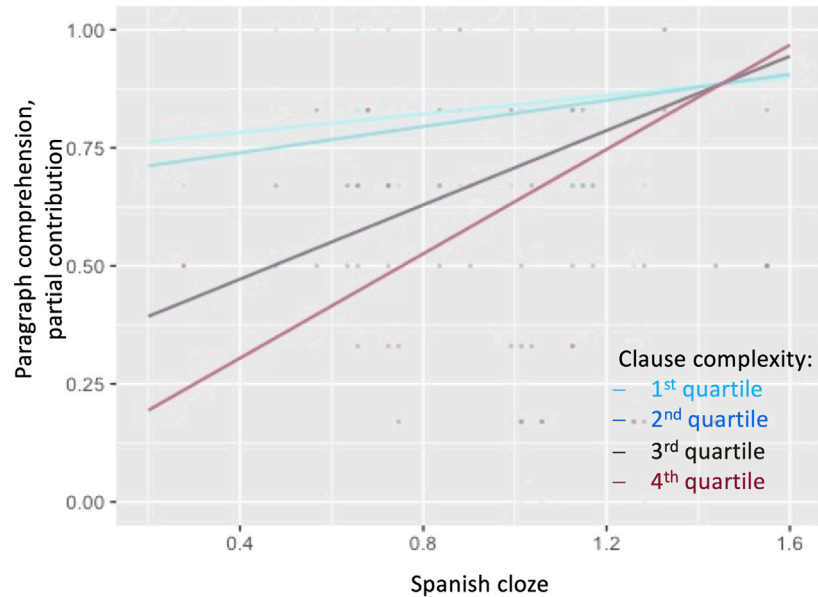


**FIGURE 3 |** Model 2.3 displaying the relationship of L1 Spanish cloze skills with paragraph comprehension for different quartiles of clause complexity ( $\beta = 0.19, p < 0.001$ ).

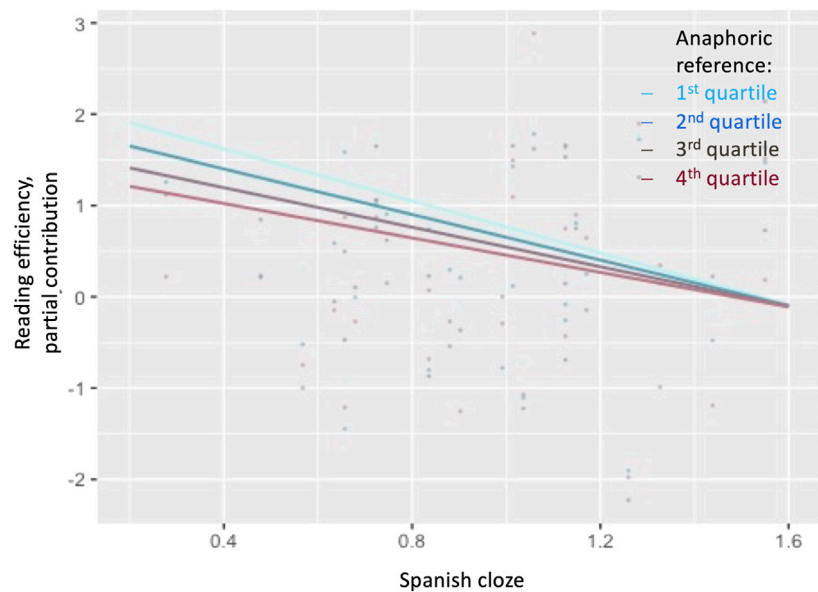


**FIGURE 4 |** Model 4.3 displaying the relationship of L1 Spanish cloze skills with reading efficiency for different quartiles of anaphoric reference (ns interaction).

movement measures (first run gaze duration and total gaze duration) aligned overall with the dimension one regression findings. Lower values of the dimension one measure were therefore indicative of faster and more efficient reading, while higher values indicated slower reading with more fixations and regressions.

Because word characteristics such as length (e.g., Rayner, 2009), frequency (e.g., Rayner and Raney, 1996), and predictability (e.g., Ehrlich and Rayner, 1981) have been widely shown to impact a variety of eye movement measures, Model 3.1 tested the importance of these measures when aggregated at the paragraph level in these experimental stimuli

**TABLE 7 |** Descriptive statistics for eye-tracking measures, by group.

|                                     | Combined (n = 59)<br>m (sd) | L1 English (n = 29)<br>m (sd) | L1 Spanish (n = 30)<br>m (sd) | Absolute group<br>mean difference<br>( $p$ of Kruskal-Wallis $\chi^2$ ) |
|-------------------------------------|-----------------------------|-------------------------------|-------------------------------|-------------------------------------------------------------------------|
| <i>Local/Word level—Early</i>       |                             |                               |                               |                                                                         |
| First fixation duration (ms)        | 235.5 (29.2)                | 234.8 (312)                   | 236.1 (27.7)                  | 1.3 (0.63)                                                              |
| Gaze duration (ms)                  | 290.6 (49.7)                | 278.8 (49.7)                  | 302.0 (47.7)                  | 23.2 (0.02)*                                                            |
| 1st pass fixation count             | 1.24 (0.11)                 | 1.19 (0.08)                   | 1.29 (0.11)                   | 0.10 (<0.001)***                                                        |
| 1st pass mean saccade amplitude (°) | 6.21 (0.95)                 | 6.55 (1.09)                   | 5.88 (0.65)                   | 0.67 (0.003)**                                                          |
| 1st pass regressions out            | 0.19 (0.07)                 | 0.19 (0.06)                   | 0.20 (0.07)                   | 0.01 (0.84)                                                             |
| 1st pass word-skipping rate         | 0.32 (0.10)                 | 0.36 (0.11)                   | 0.28 (0.07)                   | 0.07 (0.009)**                                                          |
| <i>Local/Word level—Late</i>        |                             |                               |                               |                                                                         |
| Mean fixation duration (ms)         | 232.0 (27.6)                | 232.0 (29.9)                  | 232.0 (25.7)                  | .02 (0.87)                                                              |
| Total reading time (ms)             | 366.5 (119.0)               | 322.5 (89.8)                  | 409.1 (129.3)                 | 86.6 (0.006)**                                                          |
| Mean fixation count                 | 1.57 (0.45)                 | 1.38 (0.27)                   | 1.76 (0.51)                   | 0.38 (0.001)**                                                          |
| Total regressions out               | 0.33 (0.13)                 | 0.33 (.13)                    | 0.33 (0.13)                   | 0.00 (0.96)                                                             |
| Mean saccade amplitude (°)          | 5.13 (0.10)                 | 5.51 (1.03)                   | 4.77 (0.82)                   | 0.73 (0.002)**                                                          |
| <i>Global/Text level</i>            |                             |                               |                               |                                                                         |
| Trial fixation count                | 156.51 (44.84)              | 137.14 (27.28)                | 175.24 (50.64)                | 38.10 (0.001)**                                                         |
| Trial saccade count                 | 164.17 (46.24)              | 144.05 (28.90)                | 183.63 (51.70)                | 39.57 (0.001)**                                                         |
| Words skipped (%)                   | 19.20 (7.27)                | 22.07 (6.67)                  | 16.43 (6.83)                  | 5.64 (0.001)**                                                          |
| Total trial duration (ms)           | 47655 (14404)               | 42043 (11383)                 | 53079 (15097)                 | 11036 (0.003)**                                                         |

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

**TABLE 8 |** Results of fitting a taxonomy of multi-level gaussian regression models for paragraph reading efficiency (displaying standardized coefficients).

|                                     | Model 3.1<br>$\beta$ (se) | Model 3.2<br>$\beta$ (se) | Model 4.1<br>$\beta$ (se) | Model 4.2<br>$\beta$ (se) | Model 4.3<br>$\beta$ (se) |
|-------------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Intercept                           | -0.30 (0.04)***           | -0.30 (0.04)***           | 0.35 (0.04)***            | 0.48 (0.14)**             | 0.45 (0.12)***            |
| <b>Control variables</b>            |                           |                           |                           |                           |                           |
| Group (1 = Spanish)                 | 0.60 (0.07)***            | 0.60 (0.06)***            | —                         | —                         | —                         |
| Nonverbal reasoning                 | -0.11 (0.03)***           | -0.11 (0.03)***           | -0.01 (0.04)              | -0.04 (0.04)              | —                         |
| <b>Word—level predictors</b>        |                           |                           |                           |                           |                           |
| Word predictability                 | -0.80 (0.40)*             | —                         | —                         | —                         | —                         |
| Word frequency (log)                | 0.22 (0.18)               | —                         | —                         | —                         | —                         |
| Word length                         | -1.03 (0.53)              | —                         | —                         | —                         | —                         |
| <b>Text—level predictors</b>        |                           |                           |                           |                           |                           |
| Phrase elaboration                  | —                         | -0.04 (0.05)              | -0.07 (0.08)              | —                         | —                         |
| Clausal complexity                  | —                         | -0.05 (0.06)              | -0.07 (0.08)              | —                         | —                         |
| Anaphoric reference                 | —                         | -0.08 (0.03)**            | -0.10 (0.04)*             | -0.12 (0.03)***           | -0.30 (0.10)**            |
| <b>Participant—level predictors</b> |                           |                           |                           |                           |                           |
| English word fluency                | -0.48 (0.03)***           | -0.48 (0.03)***           | -0.60 (0.05)***           | -0.64 (0.05)***           | -0.65 (0.05)***           |
| English vocabulary                  | 0.05 (0.05)               | 0.05 (0.05)               | 0.25 (0.06)***            | 0.13 (0.07)               | 0.15 (0.07)*              |
| English cloze                       | -0.13 (0.05)**            | -0.13 (0.05)**            | -0.32 (0.07)***           | -0.17 (0.07)*             | -0.20 (0.06)**            |
| Spanish word reading                | —                         | —                         | —                         | -0.07 (0.16)              | —                         |
| Spanish vocabulary                  | —                         | —                         | —                         | 1.13 (0.24)***            | 1.06 (0.22)***            |
| Spanish cloze                       | —                         | —                         | —                         | -1.18 (0.19)***           | -1.17 (0.19)***           |
| Anaphoric reference X sp. cloze     | —                         | —                         | —                         | —                         | 0.18 (0.10) <sup>†</sup>  |
| $N_{ID}$                            | 59                        | 59                        | 30                        | 30                        | 30                        |
| Variance <sub>ID</sub>              | 0.35                      | 0.35                      | 0.36                      | 0.32                      | 0.32                      |
| Observations                        | 235                       | 235                       | 119                       | 119                       | 119                       |
| AIC                                 | 307.1                     | 307.1                     | 157.5                     | 160.2                     | 155.1                     |
| BIC                                 | 524.6                     | 524.6                     | 257.7                     | 262.3                     | 254.6                     |

<sup>†</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .

(Table 7). Behavioral and syntactic predictors for the taxonomy in Table 8 were constructed in the same way as for the preceding analysis in Table 5. In Model 3.1, only average word predictability was significantly associated with reading efficiency ( $\beta = -0.80$ ,  $p = 0.04$ ); however, variation accounted

for by word predictability was colinear with syntactic complexity measures and was therefore not modeled in the remainder of the taxonomy. L1 Spanish speakers, on average, demonstrated reading efficiency outcomes 0.60 of a standard deviation higher than that of L1 English speakers ( $\beta = 0.60$ ,  $p < 0.001$ )

in Model 3.1 when controlling for nonverbal reasoning, English word fluency and English vocabulary. Unlike models in the paragraph comprehension, English word fluency in Model 3.1 was a significant predictor of reading efficiency ( $\beta = -0.48$ ,  $p < 0.001$ ). However, English vocabulary was not ( $\beta = 0.05$ ,  $p = 0.33$ ), and only English word fluency was associated with lower values of reading efficiency (i.e., shorter gaze durations, faster reading times, fewer saccades, and longer saccade amplitudes) when controlling for L1 and for nonverbal reasoning.

Model 3.2 next determined whether the syntactic predictors of paragraph comprehension were also related to reading efficiency. In this model, neither phrase ( $\beta = -0.04$ ,  $p = 0.45$ ) nor clause ( $\beta = -0.04$ ,  $p = 0.39$ ) complexity predicted reading efficiency; however, on average in the full sample, anaphoric reference (i.e., the proportion of pronouns and demonstratives, and the ratio of pronouns to nouns) was associated with better reading efficiency ( $\beta = -0.08$ ,  $p = 0.006$ ). When re-examined in Model 4.1 with L1 Spanish speakers only, results mirrored those from the full sample, with anaphoric reference similarly associated with better reading efficiency in L2 ( $\beta = -0.10$ ,  $p = 0.01$ ).

As in the paragraph response taxonomy, Model 4.2 next determined whether Spanish language and reading skills additionally contributed to explaining variance in the outcome measure. When Spanish predictors were entered into a new model, the coefficient for Spanish vocabulary was once again positive ( $\beta = 1.13$ ,  $p < 0.001$ ) such that a larger Spanish vocabulary predicted worse reading efficiency (i.e., longer fixation durations, longer reading times, shorter saccades and less word skipping). In contrast, Spanish cloze in Model 4.2 displayed a negative coefficient ( $\beta = -1.20$ ,  $p < 0.001$ ), indicating that higher standardized scores in Spanish cloze were associated with shorter fixations, faster reading times, longer saccades, and more word skipping. Furthermore, in the final model, Model 4.3 illustrated in **Figure 4**, there was a trending, but non-significant, positive interaction between the anaphoric reference measure and Spanish cloze ( $\beta = 0.18$ ,  $p = 0.08$ ), such that in passages with more anaphoric references, higher levels of Spanish cloze were trending toward, but not significantly associated with, lower decrements in reading efficiency, as compared to passages with fewer anaphoric references.

## DISCUSSION

This study examined the contributions of cross-linguistic sentence integration skills as well as the lexical skills of decoding and vocabulary knowledge to the processing and comprehension of naturalistic English texts in adolescents with either English or Spanish as their L1. In particular, it focused on the modulation of these skills by syntactic structures that have been demonstrated to challenge middle-school readers, namely, complex noun phrases, complex clauses, and anaphoric references. Middle-schoolers who spoke English or Spanish as an L1 read nonfiction passages in English that were leveled from grades 5 to 8 while eye-tracking measures were collected to assess reading efficiency (fluency and speed of reading). Participants also answered post-reading questions about the passages as a measure of comprehension.

## RQ1: Are Greater L1 and L2 Proficiency in Sentence Integration Associated With Higher Levels of L2 Text Reading Efficiency and Comprehension?

We hypothesized that greater L1 and L2 proficiency in sentence integration would be associated with higher levels of L2 text comprehension and that phrase complexity, clause complexity, and anaphoric references would all negatively modulate the expected positive effect of cross-linguistic sentence integration skills on offline text comprehension. English sentence integration and vocabulary scores were positive predictors of English text comprehension, and this relationship was similar for both L1 English and L1 Spanish speakers. This result is consistent with prior studies of English learners' reading development (e.g., Lesaux et al., 2007; Lesaux and Harris, 2017) and with most reading comprehension models, including the RSF (Perfetti and Stafura, 2014), which generally highlight the important role of skills in the language of the text in both reading fluency and comprehension.

We expected that L1 sentence integration abilities would support L2 text comprehension, and this was indeed the case when controlling for the effect of Spanish L1 vocabulary. Prior literature has found weak or nonexistent associations of L1 sentence integration with L2 reading among students in L2 dominant school environments but more consistent positive associations among students in bilingual education settings where students receive academic instruction in both languages. Although the L1 Spanish speakers in our sample, except for two in bilingual English-Spanish schools, attended almost entirely English-speaking schools, they still retained relatively balanced or Spanish-dominant language skills when we tested them in middle school. Our biliterate sample was also reported by parents to consistently engage in reading in Spanish at home, with an average of 34% of daily reading reported to take place in Spanish. Thus, in spite of educational settings similar to those found in prior studies with no CLI effects in text reading, our sample possessed proficiency in L1 Spanish reading that may have led to outcomes more similar to students in bilingual education than ones in English-only instruction. If the transfer of a common underlying proficiency, as in Cummins' (1979) proposal, requires some minimal level of L1 abilities, the home language and literacy experiences in our sample appear to have sustained at least a level of Spanish proficiency that supported the transfer of sentence integration skills to English reading.

The positive association of L1 sentence integration skills with reading was not limited to offline, post-reading comprehension but also seen in reading efficiency, where higher levels of Spanish sentence integration were associated with faster and more efficient online text processing. The parallel findings in both online behavior (seen in eye-tracking) and offline behavior (seen in the post-reading comprehension task) suggests that CLI of L1 Spanish sentence integration is not, or is not solely, a post-reading process of reasoning or strategizing about the comprehension questions asked, or re-creating a representation of the text meaning in relation to the comprehension questions asked of students after reading. Instead, these triangulated results suggest that CLI occurs during the reading process as readers are

integrating words into larger sentence and text representations. For monolingual models of text processing and comprehension such as the RSF, these findings indicate that linguistic resources beyond the language of the text can support this integration process even when lexical knowledge in that language, such as reflected in our L1 Spanish speakers' English vocabulary scores, may not provide the high-quality lexical representations of meaning that are called for in the RSF to support sentence- and text-level meaning integration. While sentence integration in the RSF model is generally assumed to be a within-language skill supporting word-to-text integration, results from both the behavioral and eye-movement analyses thus indicate that a cross-language sentence integration competency beyond vocabulary knowledge may also support text processing and comprehension in an L2.

## **RQ2: Does Syntactic Complexity of the Text Modulate the Association of L1 Sentence Integration Skills With L2 Text Reading Efficiency and Comprehension?**

Because syntactically complex texts present greater difficulty in both L1 and L2 reading, we had hypothesized that syntactic complexity of the text would modulate the association between L1 sentence integration skills and both text processing and comprehension. Our results demonstrate a significant negative main effect of syntactic complexity on paragraph comprehension. In other words, more complex texts, in terms of clause complexity, were more difficult to comprehend both for the full sample and for L1 Spanish speakers when examined separately. Further, clause complexity positively moderated the relationship of Spanish sentence integration skills with offline comprehension such that the positive association of Spanish sentence integration with L2 comprehension was more pronounced in texts with greater clause complexity. In the least complex passages, there was no evident relationship between L1 sentence integration and paragraph comprehension while in the most complex passages, higher L1 sentence integration skills were associated with better paragraph comprehension. L2 comprehension thus appears to draw more heavily on sentence integration skills developed in the L1 when sentences are more structurally complex and difficult to understand.

In online text processing, we again found that syntactic characteristics of the text were associated with reading efficiency. However, unlike for comprehension, better reading efficiency was predicted by the greater presence of anaphoric references and not by clause complexity. In other words, passages with more anaphoric references were read faster and more efficiently than those with fewer such references. In the study passages, these text features may have provided links among concepts in the text that led to more fluent and efficient online text processing. We also found a trend, but non-significant association for the interaction of anaphoric references and L1 sentence integration skills, raising the question of whether in a larger sample or a longer paradigm, these text characteristics would modulate the

facilitative main effect of L1 sentence integration on text processing efficiency.

While the present exploratory study cannot identify mechanisms for CLI in complex text, the RSF was formulated as a bottom-up, memory-based integration model, suggesting that beyond lexical representation, individual differences in short-term and working memory capacities could explain comprehension outcomes. Results of the present study do not support a memory-based explanation of individual differences as our measures of sentence integration in both Spanish and English were written and untimed, and in group comparisons, L1 Spanish and L1 English speakers were matched on verbal working memory and nonverbal reasoning. However, the study findings do suggest that in comprehending text with complex clauses that may be likely to tax memory resources, bilinguals may be drawing upon expectation-based meaning construction skills measured by our sentence integration task. These skills may be specific to features shared by Spanish and English or shared in the academic register. For example, while Spanish and English share a canonical subject-verb-object word order in simple sentences, we might speculate that the relative flexibility of Spanish regarding word order could facilitate comprehension of variation in word order in academic language marked by complex clauses. For Spanish speakers, a bias toward disregarding canonical word order in complex clauses, or toward attending to lexical cues in anaphoric references, could conceivably thus facilitate text processing and comprehension in texts containing anaphors and complex clausal structures. The crosslinguistic influence as well as the modulatory effect of text complexity on CLI demonstrated here may thus be specific to a feature such as word order, or combinatorial similarities in a particular language, Spanish, or pair of languages, Spanish and English. Alternatively, crosslinguistic influence might arise from language-general factors such as general skills with statistical learning, predictive processing, parsing, or inferencing. Future studies that employ more than one task, or more specific tasks, may disentangle these or other possibilities.

## **CONCLUSION**

Our findings extend monolingual reading models such as the RSF by providing evidence at one level, that of word to sentence and text integration, of the cross-linguistic influence of L1 sentence integration skills in L2 reading. Further, the cross-linguistic resources reflected in bilingual text processing and comprehension are dynamic and modulated by syntactic complexity of the text. Bilingual resources in integrating word meanings may thus be sensitive to text-level features not seen in studies of single word and sentence processing. The present study thus also extends word level models in bilingual research which propose a unified model of language processing and extend the implications of lexical models to sentence integration in the comprehension of connected text. Results also suggest that sentence integration may involve unified components beyond the lexicon. Future research may thus investigate how monolingual construction-integration models of reading comprehension, such as

the RSF, can incorporate bilingual processing and cross-linguistic influence, not only at the lower levels of phonology, orthography, morphology and the lexicon, but also in integrative processes that are involved in comprehending text features such as the complex clauses and anaphoric references examined in this study.

Because the study included only one global measure of sentence integration, a cloze task, one limitation is that it cannot differentiate among multiple forms of sentence integration that may have contributed to the results, such as word order sequencing, parsing, or referential association. In addition, there was substantial heterogeneity among the L1 Spanish speakers in terms of age of English acquisition and Spanish language ability. This heterogeneity reflects the linguistic and demographic mix in many United States schools and was partially accounted for using mixed effects modeling instead of simple group comparisons. However, it also demands caution in interpreting the effects of predictors that are correlated in the sample, such as English and Spanish vocabulary. Future research might attend to age of acquisition and language skill differences in a larger and longitudinal sample of L2 readers.

In summary, the current study suggests that L1 sentence integration skills can facilitate L2 reading efficiency and comprehension. L1 sentence integration skill in particular appears to support comprehension of complex clauses, even when that text is in an L2. It adds to prior educational research on cross-linguistic influence on reading outcomes using a sample representative of United States public schools, which are largely English-only and do not provide home language literacy instruction. In spite of the English-only educational background of our sample, we found a facilitatory CLI similar to that found in prior studies of children in bilingual immersion education settings, suggesting that when home language skills are developed and maintained outside of the school context, these skills can help support efficiency and comprehension of L2 reading tasks required of children at school. In supplement to bilingual education, practices and policies which support family—or community-based, out-of-school, home language development may thus also support minority language speakers' success in L2 education.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## REFERENCES

- Aguilar, G., Uccelli, P., and Galloway, E. P. (2020). Toward Bilinguality: Unpacking the Contribution of Mid-Adolescent Dual Language Learners' Spanish and English Academic Language Skills to English Reading Comprehension. *TESOL Q.* 54 (4), 1010–1036. doi:10.1002/tesq.570
- Alonso, R. A. (2019). "Crosslinguistic Influence in Second Language Acquisition," in *The Encyclopedia of Applied Linguistics*. Editor C. A. Chapelle, 1–7. doi:10.1002/9781405198431.wbeal0292 (Ed.)

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Harvard University Committee on the Use of Human Subjects (IRB16-0866). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

SL contributed to conceptualization, data collection and curation, designed and conducted the statistical analysis, and wrote the manuscript. VW contributed to conceptualization, data curation, and reviewed and edited the manuscript. LM contributed to conceptualization, data collection, and reviewed and edited the manuscript. GL contributed to conceptualization, funding acquisition, reviewed and edited the manuscript, and supervised the project. All authors contributed to the article and approved the submitted version.

## FUNDING

We would like to thank Dr. James Ryan, former Dean of the Harvard Graduate School of Education, as well as the Natural Sciences and Engineering Research Council of Canada (RGPIN-2020-05052), for providing financial support for the study (both awarded to GL). The study was carried out at the Harvard Center for Brain Science and involved the use of instrumentation supported by the NIH Shared Instrumentation Grant Program (S10OD020039).

## ACKNOWLEDGMENTS

We would like to express our sincerest gratitude to the children and families who participated in the research. We also thank the dedicated BEE lab members who assisted in recruiting and interacting with the study participants.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.651769/full#supplementary-material>

- Anderson, J. A. E., Mak, L., Keyvani Chahi, A., and Bialystok, E. (2018). The Language and Social Background Questionnaire: Assessing Degree of Bilingualism in a Diverse Population. *Behav. Res.* 50 (1), 250–263. doi:10.3758/s13428-017-0867-9
- Auerbach, J. G., Zilberman-Hayun, Y., Berger, A., and Atzaba-Poria, N. (2019). Longitudinal Pathways to reading Achievement at Age 7 years: Child and Environmental Influences. *Reading Psychol.* 40 (3), 269–292. doi:10.1080/02702711.2019.1614128
- Baayen, H., Vasishth, S., Kliegl, R., and Bates, D. (2017). The Cave of Shadows: Addressing the Human Factor with Generalized Additive Mixed Models. *J. Mem. Lang.* 94, 206–234. doi:10.1016/j.jml.2016.11.006

- Babayigit, S., and Shapiro, L. (2020). Component Skills that Underpin Listening Comprehension and reading Comprehension in Learners with English as First and Additional Language. *J. Res. Reading* 43 (1), 78–97.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67 (1), 1–48. doi:10.18637/jss.v067.i01
- Bialystok, E., Luk, G., and Kwan, E. (2005). Bilingualism, Biliteracy, and Learning to Read: Interactions Among Languages and Writing Systems. *Scientific Stud. Reading* 9 (1), 43–61. doi:10.1207/s1532799xssr0901\_4
- Bieber, G., Hulac, D. M., and Schweinle, W. (2015). An Independent Evaluation of the Technical Features of the Basic Reading Inventory. *J. Psychoeducational Assess.* 33 (3), 199–209. doi:10.1177/0734282914550808
- Blythe, H. I., and Joseph, H. S. S. L. (2011). “Children’s Eye Movements During Reading” in *The Oxford Handbook of Eye Movements*. Oxford University Press, 643–662.
- Borg, I., Groenen, P. J., and Mair, P. (2012). *Applied Multidimensional Scaling*. Springer.
- Bosch, L. J., Segers, E., and Verhoeven, L. (2020). First and Second Language Vocabulary Affect Early Second Language reading Comprehension Development. *J. Res. Reading* 43 (3), 290–308. doi:10.1111/1467-9817.12304
- Brimo, D., Apel, K., and Fountain, T. (2017). Examining the Contributions of Syntactic Awareness and Syntactic Knowledge to Reading Comprehension. *J. Res. Read* 40(1), 57–74. doi:10.1111/1467-9817.12050
- Brysbaert, M., and Biemiller, A. (2017). Test-based Age-Of-Acquisition Norms for 44 Thousand English Word Meanings. *Behav. Res.* 49 (4), 1520–1523. doi:10.3758/s13428-016-0811-4
- Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behav. Res. Methods* 41 (4), 977–990. doi:10.3758/BRM.41.4.977
- Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., and Howes, C. (2002). Development of Academic Skills from Preschool through Second Grade: Family and Classroom Predictors of Developmental Trajectories. *J. Sch. Psychol.* 40 (5), 415–436. doi:10.1016/S0022-4405(02)00107-3
- Cain, K., and Oakhill, J. (2004). “Reading Comprehension Difficulties,” in *Handbook of Children’s Literacy*. Editors T. Nunes and P. Bryant (Springer Netherlands), 313–338. doi:10.1007/978-94-017-1731-1\_18
- Canty, A. J. (2002). Resampling Methods in R: the Boot Package. *The Newsl. R Project* 2, 3. Available at: [http://cran.fhcr.org/doc/Rnews/Rnews\\_2002-3.pdf](http://cran.fhcr.org/doc/Rnews/Rnews_2002-3.pdf).
- Chung, S. C., Chen, X., and Geva, E. (2019). Deconstructing and Reconstructing Cross-Language Transfer in Bilingual reading Development: An Interactive Framework. *J. Neurolinguist.* 50, 149–161. doi:10.1016/j.jneuroling.2018.01.003
- Clark, M. K., and Kamhi, A. G. (2014). Influence of Prior Knowledge and Interest on Fourth- and Fifth-Grade Passage Comprehension on the Qualitative Reading Inventory-4. *Lshss* 45 (4), 291–301. doi:10.1044/2014\_LSHSS-13-0074
- Clifton, C., Staub, A., and Rayner, K. (2007). “Eye Movements in Reading Words and Sentences” in *Eye Movements*. Elsevier, 341–371.
- Collins, A. A., Lindström, E. R., and Compton, D. L. (2018). Comparing Students with and without reading Difficulties on reading Comprehension Assessments: A Meta-Analysis. *J. Learn. Disabil.* 51 (2), 108–123. doi:10.1177/0022219417704636
- Cop, U., Keuleers, E., Drieghe, D., and Duyck, W. (2015). Frequency Effects in Monolingual and Bilingual Natural reading. *Psychon. Bull. Rev.* 22 (5), 1216–1234. doi:10.3758/s13423-015-0819-2
- Crossley, S. A., Kyle, K., and Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating Semantic Similarity and Text Overlap. *Behav. Res.* 51 (1), 14–27. doi:10.3758/s13428-018-1142-4
- Cummins, J. (1979). *Cognitive/academic Language Proficiency, Linguistic Interdependence, the Optimum Age Question and Some Other Matters*. Ontario Institute for Studies in Education.
- Cummins, J. (2000). *Language, Power, and Pedagogy: Bilingual Children in the Crossfire*. Multilingual Matters. doi:10.21832/9781853596773
- Cunnings, I. (2017). Parsing and Working Memory in Bilingual Sentence Processing. *Biling.: Lang. Cogn.* 20 (04), 659–678. doi:10.1017/S1366728916000675
- Cutting, L. E., and Scarborough, H. S. (2006). Prediction of reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured. *Scientific Stud. Reading* 10 (3), 277–299. doi:10.1207/s1532799xssr1003\_5
- De Groot, H. (2018). *An Eye-Tracking Investigation of L2 Readers of a Novel by Agatha Christie*. dissertation. Ghent University. doi:10.21468/scipost.report.674
- de Leeuw, L., Segers, E., and Verhoeven, L. (2016). The Effect of Student-Related and Text-Related Characteristics on Student’s Reading Behaviour and Text Comprehension: An Eye Movement Study. *Scientific Stud. Reading* 20 (3), 248–263. doi:10.1080/10888438.2016.1146285
- Deacon, S. H., and Kieffer, M. (2018). Understanding How Syntactic Awareness Contributes to reading Comprehension: Evidence from Mediation and Longitudinal Models. *J. Educ. Psychol.* 110 (1), 72–86. doi:10.1037/edu0000198
- Durgunoglu, A. Y., Nagy, W. E., and Hancin-Bhatt, B. J. (1993). Cross-language Transfer of Phonological Awareness. *J. Educ. Psychol.* 85 (3), 453.
- Dussias, P. E. (2010). Uses of Eye-Tracking Data in Second Language Sentence Processing Research. *Ann. Rev. Appl. Linguist* 30, 149–166. doi:10.1017/S026719051000005X
- Ehrlich, S. F., and Rayner, K. (1981). Contextual Effects on Word Perception and Eye Movements during reading. *J. Verbal Learn. Verbal Behav.* 20 (6), 641–655. doi:10.1016/s0022-5371(81)90220-6
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., and Iglesias, A. (2006). Measures of reading Comprehension: A Latent Variable Analysis of the Diagnostic Assessment of reading Comprehension. *Scientific Stud. Reading* 10 (3), 301–322. doi:10.1207/s1532799xssr1003\_6
- Gabriele, A., Troseth, E., Martohardjono, G., and Otheguy, R. (2009). Emergent Literacy Skills in Bilingual Children: Evidence for the Role of L1 Syntactic Comprehension. *Int. J. Bilingual Education Bilingualism* 12 (5), 533–547. doi:10.1080/13670050802331816
- Garrison-Fletcher, L., Martohardjono, G., and Chodorow, M. (2019). Impact of home Language reading Skills’ on L2 reading Comprehension: a Study of Newcomer Emergent Bilinguals. *Int. J. Bilingual Education Bilingualism* 1, 1–14. doi:10.1080/13670050.2019.1671307
- Gellert, A. S., and Elbro, C. (2013). Cloze Tests May Be Quick, but Are They Dirty? Development and Preliminary Validation of a Cloze Test of reading Comprehension. *J. Psychoeducational Assess.* 31 (1), 16–28. doi:10.1177/0734282912451971
- Genesee, F., and Geva, E. (2006). “Cross-Linguistic Relationships in Working Memory, Phonological Processes, and Oral Language,” in *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (Lawrence Erlbaum Associates Publishers), 175–183.
- Geva, E., and Farnia, F. (2012). Developmental Changes in the Nature of Language Proficiency and reading Fluency Paint a More Complex View of reading Comprehension in ELL and EL1. *Read. Writ* 25 (8), 1819–1845. doi:10.1007/s11145-011-9333-8
- Geva, E., and Siegel, L. S. (2000). Orthographic and Cognitive Factors in the Concurrent Development of Basic Reading Skills in Two Languages. *Read Writ.* 12 (1), 1–30. doi:10.1023/A:1008017710115
- Godfroid, A. (2019). *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. Routledge. doi:10.4324/9781315775616
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More Use Almost Always Means a Smaller Frequency Effect: Aging, Bilingualism, and the Weaker Links Hypothesis☆. *J. Mem. Lang.* 58 (3), 787–814. doi:10.1016/j.jml.2007.07.001
- Good, R., and Kaminsky, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills*. Institute for the Development of Educational Achievement. <https://dibels.uoregon.edu/>.
- Gottardo, A., Mirza, A., Koh, P. W., Ferreira, A., and Javier, C. (2018). Unpacking Listening Comprehension: The Role of Vocabulary, Morphological Awareness, and Syntactic Knowledge in reading Comprehension. *Read. Writ* 31 (8), 1741–1764. doi:10.1007/s11145-017-9736-2
- Hoff, E. (2013). Interpreting the Early Language Trajectories of Children from Low-Ses and Language Minority Homes: Implications for Closing Achievement Gaps. *Developmental Psychol.* 49 (1), 4–14. doi:10.1037/a0027238

- Jeon, E. H., and Yamashita, J. (2014). L2 reading Comprehension and its Correlates: A Meta-Analysis. *Lang. Learn.* 64 (1), 160–212. doi:10.1111/lang.12034
- Johns, J. L. (2008). *Basic reading Inventory: Pre-primer through Grade Twelve and Early Literacy Assessments*. 10th edn. Kendall Hunt.
- Joseph, H. S. S. L., Bremner, G., Liversedge, S. P., and Nation, K. (2015). Working Memory, reading Ability and the Effects of Distance and Typicality on Anaphor Resolution in Children. *J. Cogn. Psychol.* 27 (5), 622–639. doi:10.1080/20445911.2015.1005095
- Kaufman, A. S., and Kaufman, N. L. (2004). *Manual for the Kaufman Brief Intelligence Test*. 2nd Edn. Circle Pines, MN: American Guidance Service.
- Keenan, J. M., Betjemann, R. S., and Olson, R. K. (2008). Reading Comprehension Tests Vary in the Skills They Assess: Differential Dependence on Decoding and Oral Comprehension. *Scientific Stud. Reading* 12 (3), 281–300. doi:10.1080/1088430802132279
- Keenan, J. M., and Meenan, C. E. (2014). Test Differences in Diagnosing reading Comprehension Deficits. *J. Learn. Disabil.* 47 (2), 125–135. doi:10.1177/0022219412439326
- Kieffer, M. J. (2012). Early Oral Language and Later reading Development in Spanish-speaking English Language Learners: Evidence from a Nine-Year Longitudinal Study. *J. Appl. Developmental Psychol.* 33 (3), 146–157. doi:10.1016/j.appdev.2012.02.003
- Kim, M., Crossley, S. A., and Skalicky, S. (2018). Effects of Lexical Features, Textual Properties, and Individual Differences on Word Processing Times during Second Language reading Comprehension. *Read. Writ* 31 (5), 1155–1180. doi:10.1007/s11145-018-9833-x
- Kintsch, W., and van Dijk, T. A. (1978). Toward a Model of Text Comprehension and Production. *Psychol. Rev.* 85 (5), 363–394. doi:10.1037/0033-295X.85.5.363
- Kremin, L. V., Arredondo, M. M., Hsu, L. S.-J., Satterfield, T., and Kovelman, I. (2019). The Effects of Spanish Heritage Language Literacy on English reading for Spanish-English Bilingual Children in the US. *Int. J. Bilingual Education Bilingualism* 22 (2), 192–206. doi:10.1080/13670050.2016.1239692
- Kuperman, V., Matsuki, K., and Van Dyke, J. A. (2018). Contributions of Reader- and Text-Level Characteristics to Eye-Movement Patterns during Passage reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 44 (11), 1687–1713. doi:10.1037/xlm0000547
- Kyle, K., and Crossley, S. A. (2018). Measuring Syntactic Complexity in L2 Writing Using fine-grained Clausal and Phrasal Indices. *Mod. Lang. J.* 102 (2), 333–349. doi:10.1111/modl.12468
- Leider, C. M., Proctor, C. P., Silverman, R. D., and Harring, J. R. (2013). Examining the Role of Vocabulary Depth, Cross-Linguistic Transfer, and Types of reading Measures on the reading Comprehension of Latino Bilinguals in Elementary School. *Read. Writ* 26 (9), 1459–1485. doi:10.1007/s11145-013-9427-6
- Lervåg, A., Hulme, C., and Melby-Lervåg, M. (2018). Unpicking the Developmental Relationship Between Oral Language Skills and Reading Comprehension: It's Simple, But Complex. *Child Dev.* 89 (5), 1821–1838. doi:10.1111/cdev.12861
- Lesaux, N. K., Crosson, A. C., Kieffer, M. J., and Pierce, M. (2010). Uneven Profiles: Language Minority Learners' Word reading, Vocabulary, and reading Comprehension Skills. *J. Appl. Developmental Psychol.* 31 (6), 475–483. doi:10.1016/j.appdev.2010.09.004
- Lesaux, N. K., and Harris, J. R. (2017). An Investigation of Comprehension Processes Among Adolescent English Learners with reading Difficulties. *Top. Lang. Disord.* 37 (2), 182–203. doi:10.1097/tld.0000000000000120
- Lesaux, N. K., Lipka, O., and Siegel, L. S. (2006). Investigating Cognitive and Linguistic Abilities that Influence the reading Comprehension Skills of Children from Diverse Linguistic Backgrounds. *Read. Writ* 19 (1), 99–131. doi:10.1007/s11145-005-4713-6
- Lesaux, N. K., Rupp, A. A., and Siegel, L. S. (2007). Growth in reading Skills of Children from Diverse Linguistic Backgrounds: Findings from a 5-year Longitudinal Study. *J. Educ. Psychol.* 99 (4), 821–834. doi:10.1037/0022-0663.99.4.821
- Lesaux, N. K., and Siegel, L. S. (2003). The Development of reading in Children Who Speak English as a Second Language. *Developmental Psychol.* 39 (6), 1005–1019. doi:10.1037/0012-1649.39.6.1005
- Levy, R. (2008). Expectation-based Syntactic Comprehension. *Cognition* 106 (3), 1126–1177. doi:10.1016/j.cognition.2007.05.006
- Lindsey, K. A., Manis, F. R., and Bailey, C. E. (2003). Prediction of First-Grade reading in Spanish-speaking English-language Learners. *J. Educ. Psychol.* 95 (3), 482–494. doi:10.1037/0022-0663.95.3.482
- Low, P. B., and Siegel, L. S. (2005). A Comparison of the Cognitive Processes Underlying reading Comprehension in Native English and ESL Speakers. *WL&L* 8 (2), 131–155. doi:10.1075/wll.8.2.09low
- Lu, X. (2010). Automatic Analysis of Syntactic Complexity in Second Language Writing. *Ijcl* 15 (4), 474–496. doi:10.1075/ijcl.15.4.02lu
- Lubliner, S., and Hiebert, E. H. (2011). An Analysis of English-Spanish Cognates as a Source of General Academic Language. *Biling. Res. J.* 34 (1), 76–93. doi:10.1080/15235882.2011.568589
- Luk, G., and Bialystok, E. (2013). Bilingualism Is Not a Categorical Variable: Interaction between Language Proficiency and Usage. *J. Cogn. Psychol.* 25 (5), 605–621. doi:10.1080/20445911.2013.795574
- Luk, G., and Christodoulou, J. (2016). “Assessing and Understanding the Needs of Dual-Language Learners,” in *The Leading Edge of Early Childhood Education: Linking Science to Policy for a New Generation*. Editors N. Lesaux and S. Jones (Harvard Education Press), 67–90.
- Mair, P., De Leeuw, J., and Groenen, P. (2017). *Gifi: Multivariate Analysis with Optimal Scaling. R Package-Version 0.3-7/r266*.
- Mair, Patrick., and Wilcox, R. (2016). Robust Statistical Methods in R Using the Wrs2 Package. *Behav. Res. Methods*, 1–25.
- Melby-Lervåg, M., and Lervåg, A. (2011). Cross-linguistic Transfer of Oral Language, Decoding, Phonological Awareness and reading Comprehension: A Meta-Analysis of the Correlational Evidence. *J. Res. Reading* 34 (1), 114–135. doi:10.1111/j.1467-9817.2010.01477.x
- Melby-Lervåg, M., and Lervåg, A. (2014). Reading Comprehension and its Underlying Components in Second-Language Learners: A Meta-Analysis of Studies Comparing First- and Second-Language Learners. *Psychol. Bull.* 140 (2), 409–433. doi:10.1037/a0033890
- Mikolov, T., Yih, W., and Zweig, G. (2013). “Linguistic Regularities in Continuous Space Word Representations,” in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Atlanta, GA: Association for Computational Linguistics), 746–751. Available at: <https://www.aclweb.org/anthology/N13-1090>
- Mulder, E., van de Ven, M., Segers, E., Krepel, A., de Bree, E. H., de Jong, P. F., et al. (2020). Word-to-text Integration in English as a Second Language reading Comprehension. *Read. Writ* 34, 1049–1087. doi:10.1007/s11145-020-10097-3
- Nakamoto, J., Lindsey, K. A., and Manis, F. R. (2012). Development of Reading Skills from K-3 in Spanish-Speaking English Language Learners Following Three Programs of Instruction. *Read. Writ.* 25 (2), 537–567. doi:10.1007/s11145-010-9285-4
- Nation, K., and Snowling, M. J. (2000). Factors Influencing Syntactic Awareness Skills in Normal Readers and Poor Comprehenders. *Appl. Psycholinguist.* 21 (2), 229–241. doi:10.1017/S0142716400002046
- Ordóñez, C., Carlo, M., Snow, C., and McLaughlin, B. (2002). Depth and Breadth of Vocabulary in Two Languages: Which Vocabulary Skills Transfer? *J. Edu. Psychol.* 94 (4), 719–728. doi:10.1037/0022-0663.94.4.719
- Pasquarella, A., Gottardo, A., and Grant, A. (2012). Comparing Factors Related to reading Comprehension in Adolescents Who Speak English as a First (L1) or Second (L2) Language. *Scientific Stud. Reading* 16 (6), 475–503. doi:10.1080/1088438.2011.593066
- Perfetti, C., and Stafura, J. (2014). Word Knowledge in a Theory of reading Comprehension. *Scientific Stud. Reading* 18 (1), 22–37. doi:10.1080/1088438.2013.827687
- Peterson, R. L., Boada, R., McGrath, L. M., Willcutt, E. G., Olson, R. K., and Pennington, B. F. (2017). Cognitive Prediction of reading, Math, and Attention: Shared and Unique Influences. *J. Learn. Disabil.* 50 (4), 408–421. doi:10.1177/0022219415618500
- Phillips Galloway, E., Uccelli, P., Aguilar, G., and Barr, C. D. (2020). Exploring the Cross-Linguistic Contribution of Spanish and English Academic Language Skills to English Text Comprehension for Middle-Grade Dual Language Learners. *AERA Open* 6 (1), 233285841989257. doi:10.1177/2332858419892575
- Phillips Galloway, E., and Uccelli, P. (2019). Examining Developmental Relations between Core Academic Language Skills and reading Comprehension for

- English Learners and Their Peers. *J. Educ. Psychol.* 111 (1), 15–31. doi:10.1037/edu0000276
- Prevoo, M. J. L., Malda, M., Mesman, J., and van IJzendoorn, M. H. (2016). Within- and Cross-Language Relations between Oral Language Proficiency and School Outcomes in Bilingual Children with an Immigrant Background. *Rev. Educ. Res.* 86 (1), 237–276. doi:10.3102/0034654315584685
- Proctor, C. P., August, D., Snow, C., and Barr, C. D. (2010). The Interdependence Continuum: A Perspective on the Nature of Spanish-English Bilingual Reading Comprehension. *Bilingual Res. J.* 33 (1), 5–20. doi:10.1080/15235881003733209
- Proctor, C. P., Carlo, M., August, D., and Snow, C. (2005). Native Spanish-speaking Children reading in English: Toward a Model of Comprehension. *J. Educ. Psychol.* 97 (2), 246–256. doi:10.1037/0022-0663.97.2.246
- Proctor, C. P., Harring, J. R., and Silverman, R. D. (2017). Linguistic Interdependence between Spanish Language and English Language and reading: A Longitudinal Exploration from Second through Fifth Grade. *Bilingual Res. J.* 40 (4), 372–391. doi:10.1080/15235882.2017.1383949
- Proctor, C. P., Silverman, R. D., Harring, J. R., and Montecillo, C. (2012). The Role of Vocabulary Depth in Predicting reading Comprehension Among English Monolingual and Spanish-English Bilingual Children in Elementary School. *Read. Writ* 25 (7), 1635–1664. doi:10.1007/s11145-011-9336-5
- Quinn, J. M., and Wagner, R. K. (2018). Using Meta-Analytic Structural Equation Modeling to Study Developmental Change in Relations between Language and Literacy. *Child. Dev.* 89 (6), 1956–1969. doi:10.1111/cdev.13049
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*; 2015.
- Rayner, Keith. (2009). Eye Movements in reading: Models and Data. *J. Eye Movement Res.* 2 (5), 1. doi:10.16910/jemr.2.5.2
- Rayner, Keith. (2012). *Eye Movements in reading: Perceptual and Language Processes*. Elsevier.
- Rayner, K. (1998). Eye Movements in reading and Information Processing: 20 Years of Research. *Psychol. Bull.* 124 (3), 372–422. doi:10.1037/0033-2909.124.3.372
- Rayner, K., and Raney, G. E. (1996). Eye Movement Control in reading and Visual Search: Effects of Word Frequency. *Psychon. Bull. Rev.* 3 (2), 245–248. doi:10.3758/bf03212426
- Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., et al. (2013). Using E-Z Reader to Examine the Concurrent Development of Eye-Movement Control and Reading Skill. *Dev. Rev.* 33 (2), 110–149. doi:10.1016/j.dr.2013.03.001
- Relyea, J. E., and Amendum, S. J. (2020). English Reading Growth in Spanish-Speaking Bilingual Students: Moderating Effect of English Proficiency on Cross-Linguistic Influence. *Child Dev.* 91 (4), 1150–1165. doi:10.1111/cdev.13288
- Reynolds, C. R., and Voress, J. K. (2007). *TOMAL-2 Test of Memory and Learning*. Pro-Ed.& Pro-ed (Firm).
- Rigby, R. A., and Stasinopoulos, D. M. (2005). Generalized Additive Models for Location, Scale and Shape (With Discussion). *J. R. Stat. Soc C* 54, 507–554. doi:10.1111/j.1467-9876.2005.00510.x
- Rossi, E., Krass, K., and Kootstra, G. J. (2019). Psycholinguistic Methods in Multilingual Research. *The Handbook Neurosci. Multilingualism*, 75–99. doi:10.1002/9781119387725.ch4
- Scott, C. M., and Balthazar, C. H. (2010). The Grammar of Information: Challenges for Older Students with Language Impairments. *Topics in Language Disorders* 30 (4), 288–307. doi:10.1097/TLD.0b013e3181f90878
- Siu, C. T.-S., and Ho, C. S.-H. (2015). Cross-language Transfer of Syntactic Skills and reading Comprehension Among Young Cantonese-English Bilingual Students. *Read. Res. Q.* 50 (3), 313–336. doi:10.1002/rrq.101
- Siu, C. T.-S., and Ho, S.-H. (2020). A Longitudinal Investigation of Syntactic Awareness and reading Comprehension in Chinese-English Bilingual Children. *Learn. Instruction* 67, 101327. doi:10.1016/j.learninstruc.2020.101327
- Snell, J., and Theeuwes, J. (2020). A story about Statistical Learning in a story: Regularities Impact Eye Movements during Book reading. *J. Mem. Lang.* 113, 104127. doi:10.1016/j.jml.2020.104127
- Sohail, J. (2015). The within- and Cross-Language Role of Syntactic Awareness in reading Comprehension Among French Immersion Students. PhD Thesis. Available at: <http://search.proquest.com.ezp-prod1.hul.harvard.edu/docview/1753920158?accountid=11311>.
- Sorenson Duncan, T., and Paradis, J. (2020). How Does Maternal Education Influence the Linguistic Environment Supporting Bilingual Language Development in Child Second Language Learners of English?. *Int. J. Bilingualism* 24 (1), 46–61. doi:10.1177/1367006918768366
- SR Research Ltd (). *EyeLink and Experiment Builder*. Ontario, Canada: SR Research Ltd. Mississauga.
- Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., and De Bastiani, F. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*. Chapman and Hall/CRC. doi:10.1201/b21973
- Swanson, H. L., Rosston, K., Gerber, M., and Solari, E. (2008). Influence of Oral Language and Phonological Awareness on Children's Bilingual reading. *J. Sch. Psychol.* 46 (4), 413–429. doi:10.1016/j.jsp.2007.07.002
- Taylor, W. L. (1953). "Cloze Procedure": A New Tool for Measuring Readability. *Journalism Q.* 30 (4), 415–433. doi:10.1177/107769905303000401
- Tiffin-Richards, S. P., and Schroeder, S. (2018). The Development of Wrap-Up Processes in Text reading: A Study of Children's Eye Movements. *J. Exp. Psychol. Learn. Mem. Cogn.* 44 (7), 1051–1063. doi:10.1037/xlm0000506
- Torgesen, J., Wagner, R., and Rashotte, C. (2011). in *TOWRE-2*. Pro-Ed (Austin, TX: Test of Word Reading Efficiency).
- Trace, J. (2020). Clozing the gap: How Far Do Cloze Items Measure?. *Lang. Test.* 37 (2), 235–253. doi:10.1177/0265532219888617
- Tremblay, A. (2011). Proficiency Assessment Standards in Second Language Acquisition Research. *Stud. Second Lang. Acquis* 33 (3), 339–372. doi:10.1017/S0272263111000015
- Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., and Sánchez, E. (2015a). Core Academic Language Skills: An Expanded Operational Construct and a Novel Instrument to Chart School-Relevant Language Proficiency in Preadolescent and Adolescent Learners. *Appl. Psycholinguistics* 36 (5), 1077–1109. doi:10.1017/S014271641400006X
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., and Dobbs, C. L. (2015b). Beyond Vocabulary: Exploring Cross-Disciplinary Academic-Language Proficiency and its Association with reading Comprehension. *Read. Res. Q.* 50 (3), 337–356. doi:10.1002/rrq.104
- van den Bosch, L. J., Segers, E., and Verhoeven, L. (2018). Online Processing of Causal Relations in Beginning First and Second Language Readers. *Learn. Individual Differences* 61, 59–67. doi:10.1016/j.lindif.2017.11.007
- Van der Loo, M. P. J. (2014). The Stringdist Package for Approximate String Matching. *R. J.* 6 (1), 111–122. doi:10.32614/rj-2014-011
- Vanderwal, T., Eilbott, J., and Castellanos, F. X. (2019). Movies in the Magnet: Naturalistic Paradigms in Developmental Functional Neuroimaging. *Dev. Cogn. Neurosci.* 36, 100600. doi:10.1016/j.dcn.2018.10.004
- Van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press.
- Verhoeven, L., and van Leeuwe, J. (2009). Modeling the Growth of Word-Decoding Skills: Evidence from Dutch. *Scientific Stud. Reading* 13 (3), 205–223. doi:10.1080/10888430902851356
- von der Malsburg, T., and Angele, B. (2017). False Positives and Other Statistical Errors in Standard Analyses of Eye Movements in reading. *J. Mem. Lang.* 94, 119–133. doi:10.1016/j.jml.2016.10.003
- Whitford, V., and Joannis, M. F. (2018). Do eye Movements Reveal Differences between Monolingual and Bilingual Children's First-Language and Second-Language reading? A Focus on Word Frequency Effects. *J. Exp. Child Psychol.* 173, 318–337. doi:10.1016/j.jecp.2018.03.014
- Whitford, V., and Titone, D. (2015). Second-language Experience Modulates Eye Movements during First- and Second-Language Sentence reading: Evidence from a Gaze-Contingent Moving Window Paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 41 (4), 1118–1129. doi:10.1037/xlm0000093
- Whitford, V., and Titone, D. (2012). Second-language Experience Modulates First- and Second-Language Word Frequency Effects: Evidence from Eye Movement

- Measures of Natural Paragraph reading. *Psychon. Bull. Rev.* 19 (1), 73–80. doi:10.3758/s13423-011-0179-5
- Whitford, V., and Titone, D. (2014). The Effects of Reading Comprehension and Launch Site on Frequency-Predictability Interactions during Paragraph Reading. *Q. J. Exp. Psychol.* 67 (6), 1151–1165. doi:10.1080/17470218.2013.848216
- Whitford, V., and Titone, D. (2017). The Effects of Word Frequency and Word Predictability during First- and Second-Language Paragraph reading in Bilingual Older and Younger Adults. *Psychol. Aging* 32 (2), 158–177. doi:10.1037/pag0000151
- Wolf, M., and Denckla, M. B. (2005). *RAN/RAS: Rapid Automatized Naming and Rapid Alternating Stimulus Tests*. Pro-ed. Austin, TX.
- Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., and Alvarado, C. G. (2005). *Woodcock-Muñoz Language Survey-Revised*. Itasca, IL: Riverside.
- Zeileis, A., Wiel, M. A., Hornik, K., and Hothorn, T. (2008). Implementing a Class of Permutation Tests: The coin Package. *J. Stat. Softw.* 28 (8), 1–23.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Leon Guerrero, Whitford, Mesite and Luk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Do Cross-Language Script Differences Enable Bilinguals to Function Selectively When Speaking in One Language Alone?

Noriko Hoshino<sup>1\*</sup>, Anne L. Beatty-Martínez<sup>2</sup>, Christian A. Navarro-Torres<sup>3</sup> and Judith F. Kroll<sup>3</sup>

<sup>1</sup>Department of English, Tsuda University, Kodaira, Japan, <sup>2</sup>Department of Psychology, McGill University, Montréal, QC, Canada,

<sup>3</sup>Department of Language Science, University of California, Irvine, Irvine, CA, United States

## OPEN ACCESS

### Edited by:

Anat Prior,  
University of Haifa, Israel

### Reviewed by:

Chuchu Li,  
University of California, San Diego,  
United States  
Irina Elgort,  
Victoria University of Wellington,  
New Zealand

### \*Correspondence:

Noriko Hoshino  
noriko@tsuda.ac.jp

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 16 February 2021

**Accepted:** 07 June 2021

**Published:** 22 June 2021

### Citation:

Hoshino N, Beatty-Martínez AL,  
Navarro-Torres CA and Kroll JF (2021)  
Do Cross-Language Script Differences  
Enable Bilinguals to Function  
Selectively When Speaking in One  
Language Alone?.  
Front. Commun. 6:668381.  
doi: 10.3389/fcomm.2021.668381

The present study examined the role of script in bilingual speech planning by comparing the performance of same and different-script bilinguals. Spanish-English bilinguals (Experiment 1) and Japanese-English bilinguals (Experiment 2) performed a picture-word interference task in which they were asked to name a picture of an object in English, their second language, while ignoring a visual distractor word in Spanish or Japanese, their first language. Results replicated the general pattern seen in previous bilingual picture-word interference studies for the same-script, Spanish-English bilinguals but not for the different-script, Japanese-English bilinguals. Both groups showed translation facilitation, whereas only Spanish-English bilinguals demonstrated semantic interference, phonological facilitation, and phono-translation facilitation. These results suggest that when the script of the language not in use is present in the task, bilinguals appear to exploit the perceptual difference as a language cue to direct lexical access to the intended language earlier in the process of speech planning.

**Keywords:** different-script bilinguals, cross-language activation, semantic interference, phono-translation, picture-word interference task

## INTRODUCTION

Although bilinguals are able to speak each language without apparent intrusion of the other language, experimental studies demonstrate that both languages are active even when utterances are planned in one language alone and that bilinguals eventually select the intended language (see Costa, 2005; Kroll et al., 2006, Kroll et al., 2008, for reviews). A notable feature of much of the research on this topic is that it has examined the performance of bilinguals whose two languages use the same written script and are therefore potentially ambiguous with respect to language status. For speakers of languages like Dutch and English or Spanish and Catalan, there are many words that have similar orthography and phonology. Cross-language ambiguity may extend the process of language selection because ambiguity may increase cross-language competition. The question we ask in the present paper is whether differences in the written script of two languages can effectively reduce activation of the language not in use to allow bilinguals to select the target language, the language of production, earlier in the process of speech planning. If bilinguals can exploit cross-language differences to enable lexical access to be language selective, they may be better able to constrain the scope of competition and reduce functional demands on control processes. We examined this question by comparing

the performance of Spanish-English and Japanese-English bilinguals in a picture-word interference task.

Past studies on different-script word production have reported findings that are similar to those for same-script production in that the consequences of cross-language competition appear to be evident (e.g., Hoshino and Kroll, 2008; Guo et al., 2011; Misra et al., 2012; Moon and Jiang, 2012). For example, Hoshino and Kroll (2008) found that both Japanese-English and Spanish-English bilinguals produced cognate facilitation in a simple picture naming task in English, their second language (L2). In simple picture naming, the written script of each language is absent but Hoshino and Kroll hypothesized that a bilingual's experience with two languages that differ not only in script but in a variety of lexical and syntactic features might also serve as a cue to allow speech planning to be selective. Finding significant cognate facilitation for Japanese-English bilinguals means that the phonology of the Japanese name of the picture was activated when planning to speak the word in English. However, finding phonological activation of the first language (L1) during speech planning in L2 in the absence of the printed word in L1 does not mean that script information is unimportant when it is present; it is only that it does not appear to modulate processing when it is absent.

Although it might seem that the written lexical form should not influence word production, there is a great deal of evidence to suggest that semantics, phonology, and orthography are active to some degree in both comprehension and production (e.g., Tan and Perfetti, 1999; Damian and Bowers, 2003; Ziegler et al., 2003; Chéreau et al., 2007; Bi et al., 2009; Pattamadilok et al., 2009; Zhang et al., 2009; Zhang and Weekes, 2009; Rastle et al., 2011). The effects of orthography on production depend on task demands and have been reported to be reliable only when the written lexical form is overtly present in the task such as form preparation and picture-word interference paradigms (e.g., Roelofs, 2006; Alario et al., 2007; Damian and Bowers, 2009). The resonance across lexical codes suggests that each code may eventually activate whatever information is similar to it both within and across languages. For different-script languages, the implication is that the presence of phonological overlap may be sufficient to observe cross-language interactions. However, the cross-language phonological effect of the sort that Hoshino and Kroll (2008) reported does not mean that different-script bilinguals are unable to exploit language-specific cues when the words are actually present. The current study asks the question of whether different-script bilinguals can exploit script differences when the words are actually present in the task.

Models of lexical access in bilingual word production must address the locus and manner of language selection if the intention to speak one language alone is not sufficient to restrict activation to that language (for reviews, see Costa, 2005; Kroll et al., 2006; Kroll et al., 2008). Although past bilingual research concurs with the view of nonselective activation, particularly in L2 production, there has been debate over the manner of selection. At issue is whether all activated lexical alternatives become candidates for selection. According to language-nonspecific (competition for selection) models, candidates from both languages are active competitors (e.g.,

Hermans et al., 1998; Kroll et al., 2008; Hoshino and Thierry, 2011). On this view, there may be inhibition of alternatives in the unintended language (e.g., Meuter and Allport, 1999; Philipp and Koch, 2009). In contrast, language-specific models (e.g., Costa et al., 1999) assume that lexical alternatives may be active in both languages but only those in the response language are considered for selection. Kroll et al. (2008) describe this as a "mental firewall" model because the assumption is that activity on the wrong side of the firewall has little consequence for lexical selection. Alternatively, some models assume that there is no competition even within a language in the process of lexical selection. The Response Exclusion Hypothesis posits that the target and non-target alternatives compete at a post-lexical level, not at a lexical level (Mahon et al., 2007).

In theory, cross-language script differences could affect activation and/or selection. If script differences provide categorical cues to language membership, then there might not be any activation of the non-target language or that activation might subside more quickly relative to a same-script language. The results of Hoshino and Kroll (2008) suggest that in the absence of perceptual information that might cue the language of production, there appears to be similar activation of the non-target language for same and different-script bilinguals. Thus, knowledge of the different script alone does not seem to suffice to reduce the activation of cross-language phonology. Alternatively, both languages might be activated but the different script information might better enable selection to occur at an earlier stage of speech planning than otherwise possible. Again, the presence of similar cognate facilitation for both same and different-script bilinguals suggests that different language script itself does not alter the locus of language selection. However, without an explicit test of how the other language script is processed when it is perceptually present, it is premature to conclude that script does not influence cross-language activation or selection.

Debate in the past literature on bilingual language production regarding the locus and manner of selection results in part from the use of different experimental paradigms. Many bilingual production studies used the picture-word interference paradigm in which bilinguals name a picture in one language while ignoring a visually or auditorily presented distractor word in the same or other language (e.g., Hermans et al., 1998; Costa and Caramazza, 1999; Costa et al., 1999; Costa et al., 2003; Knipsky and Amrhein, 2007; Hoshino and Thierry, 2011; Boukadi et al., 2015; Giezen and Emmorey, 2016). In these experiments, the relation between the distractor word and the picture's name is varied along with the timing of the distractor presentation relative to the picture, and the language of the distractor. The logic of these studies is to examine the time course of distractor effects as a way of identifying the activity of the non-target language during each stage of production. Past research using picture-word interference has reported compelling effects of the language not in use. The evidence on picture-word interference with same-script bilinguals suggests that even when bilinguals are producing words in one language alone and the distractor is in the same language as production, there is momentary activation of the other language (e.g.,

Hermans et al., 1998; Hoshino and Thierry, 2011). In other words, the non-target language is activated in single language as well as mixed language contexts.

Although most of the past research using the picture-word interference paradigm has examined the performance of bilinguals whose languages share the same Roman alphabets, a few studies have examined the effect of language similarity on cross-language activation and lexical selection. Boukadi et al. (2015) replicated past findings for same-script bilinguals in a group of different-script Arabic-French bilinguals, with the distractor presented auditorily in the non-target language. However, no cross-language effect was observed when the language of the distractor was the same as the target language, the language of production. Similarly, Giezen and Emmorey (2016) showed that hearing bilinguals who were proficient users of English and American Sign Language (ASL), activated English when naming pictures in ASL. One noticeable finding of this study was that unlike past picture-word interference research with same-script bilinguals (e.g., Hermans et al., 1998; Costa et al., 1999; Hoshino and Thierry, 2011) and different-script bilinguals (Boukadi et al., 2015), no semantic interference effect was observed for bimodal bilinguals when they ignored semantically related English words while producing ASL. Giezen and Emmorey (2016) suggested that semantically related English distractor words and target ASL signs do not compete in the post-lexical articulatory buffer or that there might be time course differences between sign production and spoken word production. Although bimodal bilinguals can produce two languages (both signs and words) simultaneously unlike unimodal bilinguals, the control mechanism involved in the process of speech production is more similar than different (e.g., Blanco-Elorrieta et al., 2018; Emmorey et al., 2020a). The important point of these past studies with different-script and bimodal bilinguals is that there was cross-language activation when distractor words were presented auditorily in the non-target language. It is possible that different scripts would modulate cross-language activation if the distractor words were presented visually rather than auditorily.

What evidence exists that visually presented distractors might provide cues to the language of production? Miller and Kroll (2002) reported evidence for the language cue hypothesis in a translation Stroop study with same-script bilinguals. The translation Stroop task is formally similar to picture-word interference but instead of a picture as the initiating event, a word is presented for translation. The task is to translate the word as quickly as possible while ignoring a distractor word. When the distractor word appeared in the language of production (e.g., see a word in Spanish to be translated into English and the distractor word appears in English), Miller and Kroll found semantic interference and form facilitation replicating previous translation and picture-word interference studies (La Heij et al., 1990). However, when the distractor word appeared in the language of the target word to be translated (e.g., see a word in Spanish to be translated into English but the distractor appears in Spanish), they found that there was neither semantic interference nor form facilitation. They argued that in

translation, unlike picture naming, there is a cue to language membership available in the target word that initiates speech planning. If the word appears in Spanish in a translation task, the bilingual knows not to speak Spanish. If script differences function as explicit cues to language status and if bilinguals can exploit language-specific information, then the process of planning the spoken utterance becomes similar to a within-language process in which only candidates in the language to be produced compete for selection. Because English and Japanese differ in script, there may be stronger cues for language status than for English and Spanish. Indeed, color Stroop studies with different-script bilinguals and trilinguals have demonstrated that when two or three languages differ in script, bilinguals and trilinguals experience less cross-language interference (e.g., Smith and Kirsner, 1982; Chen and Ho, 1986; Brauer, 1998; Lee and Chan, 2000; Van Heuven et al., 2011).

We note that the results of color Stroop studies are limited given that they include only color names and the number of mappings is small. Therefore, in the present study, we used a picture-word interference paradigm, which allowed us to include a variety of conditions. Similar to past bilingual studies that have used this paradigm, the experiment we report included four types of distractor words in relation to the name of the target picture: phonologically related to the picture name, semantically related to the picture name, phonologically related to the translation of the picture name (phono-translation), and the translation of the picture name. Each of these distractor words was matched with an unrelated control. Same-script Spanish-English bilinguals (Experiment 1) and different-script Japanese-English bilinguals (Experiment 2) were asked to name noncognate pictures in their L2 English while ignoring visually presented distractors in their L1 Spanish or Japanese. In other words, the present study examined picture naming in L2 in the presence of distractors in L1 because those are the conditions that typically produce the largest cross-language effects. For Spanish-English bilinguals, we expected to replicate previously reported results, i.e., phonological facilitation, semantic interference, phono-translation interference, and translation facilitation. On the other hand, if distractor words in a different-script language provide a language cue to production that bilinguals are able to exploit, there should be no effect of distractor conditions when Japanese-English bilinguals produce picture names in English in the presence of Japanese distractor words.

## EXPERIMENT 1: SPANISH-ENGLISH BILINGUALS

### Method Participants

Forty-eight Spanish-English bilinguals participated in Experiment 1. They were all living in the L2 environment at the time of testing. After completing the main picture-word interference task, participants were given a language history questionnaire (Tokowicz et al., 2004), an English lexical decision task (Azuma and Van Orden, 1997), a semantic

**TABLE 1 |** Comparisons of Spanish-English (Experiment 1) and Japanese-English bilinguals (Experiment 2) on proficiency and cognitive measures.

| Measure                             | Spanish-English<br>(Experiment 1) | Japanese-English<br>(Experiment 2) | p      |
|-------------------------------------|-----------------------------------|------------------------------------|--------|
| Age (years)                         | 25.7 (5.9)                        | 27.2 (7.7)                         | 0.32   |
| L1 self-rating (10 pt scale)        | 9.4 (0.8)                         | 9.1 (1.1)                          | 0.11   |
| L2 self-rating (10 pt scale)        | 8.5 (1.3)                         | 7.0 (1.8)                          | <0.001 |
| Daily L1 usage (%)                  | 43.2 (25.1)                       | 39.5 (26.8)                        | 0.51   |
| Daily L2 usage (%)                  | 57.9 (23.8)                       | 57.8 (26.2)                        | 0.98   |
| Age of L2 acquisition (years)       | 8.3 (4.8)                         | 10.6 (3.7)                         | 0.01   |
| Length of immersion (months)        | 87.4 (78.1)                       | 79.1 (72.1)                        | 0.62   |
| Lexical decision                    |                                   |                                    |        |
| RT for nonword(ms)                  | 937 (260)                         | 1,017 (301)                        | 0.18   |
| RT for word (ms)                    | 683 (112)                         | 729 (177)                          | 0.17   |
| Accuracy for nonword (%)            | 83.3 (11.6)                       | 78.8 (10.8)                        | 0.07   |
| Accuracy for word (%)               | 94.1 (3.7)                        | 93.5 (3.0)                         | 0.44   |
| Verbal fluency in L1 <sup>a</sup>   | 12.2 (2.9)                        | 11.8 (1.7)                         | 0.40   |
| Verbal fluency in L2 <sup>a</sup>   | 11.1 (2.0)                        | 10.8 (2.0)                         | 0.36   |
| Simon effect (ms) <sup>b</sup>      | 46.7 (33.4)                       | 28.9 (22.2)                        | 0.01   |
| RT for neutral (ms)                 | 466 (96)                          | 404 (52)                           | <0.001 |
| RT for congruent (ms)               | 446 (97)                          | 390 (58)                           | 0.002  |
| RT for incongruent (ms)             | 493 (94)                          | 419 (52)                           | <0.001 |
| Accuracy for neutral (%)            | 99.4 (1.6)                        | 98.8 (2.0)                         | 0.18   |
| Accuracy for congruent (%)          | 99.1 (1.7)                        | 98.5 (2.1)                         | 0.16   |
| Accuracy for incongruent (%)        | 96.7 (4.2)                        | 97.0 (3.4)                         | 0.74   |
| Operation span (0–60) <sup>c</sup>  | 33.9 (10.5)                       | 39.2 (7.4)                         | 0.01   |
| RT for equation judgment (ms)       | 2537 (235)                        | 2182 (251)                         | <0.001 |
| Errors for equation judgment (0–60) | 14.6 (8.3)                        | 8.6 (4.7)                          | <0.001 |

Standard deviations are in parentheses.

<sup>a</sup>The verbal fluency score is the mean exemplars per category.

<sup>b</sup>The Simon effect is the difference in RTs between congruent and incongruent conditions.

<sup>c</sup>The operation span is the number of words that were recalled correctly among correct responses to the equation judgment, which were not considered as outliers.

verbal fluency task (Linck et al., 2009), Simon task (Bialystok et al., 2004; Linck et al., 2008), and an operation span task (Tokowicz et al., 2004) as a means to match the two bilingual groups. The results of the additional tasks describing the characteristics of the participants (the language history questionnaire) and measuring their language proficiency (the lexical decision task and the semantic verbal fluency task) and cognitive abilities/resources (the Simon task and the operation span task) are summarized in **Table 1**.

## Materials

### Pictures

Sixteen black-and-white line drawings were sampled from Snodgrass and Vanderwart (1980), Székely et al. (2003) and Székely et al. (2004) based on the following criteria: 1) all pictures had noncognate names in English, Spanish, and Japanese; 2) all pictures were typically written in kanji for Japanese<sup>1</sup>; 3) the phonological onset of the English name of pictures was restricted to phonology that was shared with

Spanish and Japanese.<sup>2</sup> In addition to the experimental pictures, eight filler pictures that met the first two criteria were included in the present experiment. None of the fillers was the same as the experimental pictures.

### Distractor Words

For each of the pictures, four types of distractor words were selected in Spanish: phonologically related to the English picture name, semantically related to the English picture name, Spanish translation name of the English picture name, phonologically related to the Spanish translation name of the English picture name (phono-translation) (see **Table 2** for examples). The following criteria were used to select each type of distractor: 1) distractors that were phonologically related to the English picture name or to the Spanish translation of the English picture name were matched on phonological onset with the English picture name and were not also semantically related to the target picture; 2) semantically related distractors were largely from the same semantic categories and were not phonologically related to the English or Spanish name of the picture. Each of these related

<sup>1</sup>We included the criteria (2) and (3) to use an identical set of pictures for Experiments 1 and 2. In Japanese, some words can be written both in kanji and in kana (hiragana and katakana), whereas others can be written only in kana (hiragana and katakana). Although some words can be written in either writing system, there are preferences for one rather than the other. The preference was determined by the frequency of each form (Amano and Kondo, 2000).

<sup>2</sup>For example, “frog” was not included according to the third criterion because the consonant cluster/fr/ was not possible in Japanese. Although these criteria limited the number of items in the present experiment, it was critical to use these criteria to make the contribution of phonologically related distractors equivalent for Spanish-English and Japanese-English bilinguals.

**TABLE 2 |** Examples of distractors for the picture “envelope” by distractor type and relatedness and characteristics of Spanish distractors.

| Distractor type   | Examples           |                   | Frequency |           | Length    |           |
|-------------------|--------------------|-------------------|-----------|-----------|-----------|-----------|
|                   | Related            | Unrelated         | Related   | Unrelated | Related   | Unrelated |
| Phonological      | enchufe (plug)     | rodilla (knee)    | 2.015     | 2.004     | 5.4 (2.3) | 5.6 (2.4) |
| Semantic          | tarjeta (postcard) | alicates (pliers) | 1.735     | 1.819     | 5.8 (2.4) | 6.1 (2.6) |
| Translation       | sobre (envelope)   | hombre (man)      | 2.061     | 2.065     | 5.2 (2.2) | 5.3 (2.2) |
| Phono-translation | sobriño (nephew)   | paloma (pigeon)   | 1.614     | 1.603     | 5.8 (2.6) | 5.8 (2.6) |

(1) The translation of the distractor word is given in parentheses. (2) The number of letters is provided without parentheses and the number of syllables is provided with parentheses. Frequency was from Alameda and Cuetos (1995) and log frequency was computed based on the values from the source.

distractors had an unrelated control that was matched item-by-item based on length (number of characters and syllables) and frequency [all  $ps > 0.10$ ] (see **Table 2**). Because it was not possible to match distractors across distractor types completely without weakening the relation of the pair within each condition, we analyzed the data for each distractor type separately. Each of the filler pictures also had a distractor word in Spanish. The complete set of the experimental items is provided (**Supplementary Appendix A**).

In the present study, each experimental picture was presented eight times but with different distractors so that none of the distractor words was repeated. Likewise, filler pictures were also repeated eight times. Unlike the experimental trials, however, distractors for filler pictures were each presented four times. Each list had eight blocks and each of the eight blocks included 16 experimental pictures and eight filler pictures. There were two items for each type of distractor and each of its unrelated controls per block for the experimental pictures. Each block started with two filler trials and the critical trials and the rest of the filler trials were presented randomly within the block. The order of blocks was counterbalanced across participants.

## Procedure

Participants first received written instructions in English on the computer screen. They were informed that a series of pictures would be presented with an L1 (Spanish) distractor word one at a time on the computer screen. Their task was to name the pictured object in English as quickly and accurately as possible while ignoring the L1 distractor word. At the beginning of each trial, a fixation sign (+) was presented at the center of the computer screen. At the press of a button, the fixation sign was replaced with a blank screen and 500 ms after the offset of the fixation sign, a picture was presented. A distractor word appeared in red at the center of the picture 25 ms after the onset of the presentation of the picture. The picture and the distractor word were presented until the participants responded or for 5,000 ms. If they did not know the name of the object, they were told to say “no”. After they responded, a blank screen was presented for 500 ms and a fixation sign appeared again. The 25 ms delay was included to ensure that participants could first see the pictured object clearly. Eight practice trials were presented twice prior to the experimental trials. The pictures and distractors used in the practice trials were different from the experimental items.

## Data Trimming Procedure

Recorded picture naming responses were first transcribed and coded for accuracy. We included only the expected picture names as correct responses in order to maintain the phonological manipulation. Responses that deviated from the expected picture name, responses that started with an article or hesitation, and “no” responses were scored as errors (5.6%). Responses that the microphone did not detect were eliminated as technical errors (<0.1%). Correct responses that were less than 300 ms or greater than 2,500 ms were identified as outliers (1.1%) and excluded from the analyses.

## Data Analysis

Statistical analyses were performed using linear and generalized mixed-effects models (Baayen et al., 2008) in the lme4 software package (v. 1.1.23; Bates et al., 2015) in the R statistical software environment (v. 4.0.2; R Core Team, 2020). For picture naming latency analyses, models had log response times (RTs) as the dependent variable and distractor type (related or unrelated) as a dummy-coded fixed effect. For picture naming accuracy, we ran logistic mixed-effects regression with accuracy as the dependent variable and distractor type (related vs. unrelated) as a dummy-coded fixed effect. To guard against Type I errors and increase generalizability (Barr et al., 2013), we attempted to fit random effects including both random intercepts and random slopes. However, these models returned a warning message for singularity (i.e., where one or more variances are estimated as zero). To allow a non-singular fit, random slopes were removed and thus the random effect structure for both RT and accuracy models contained only random intercepts for participants and items.<sup>3</sup>

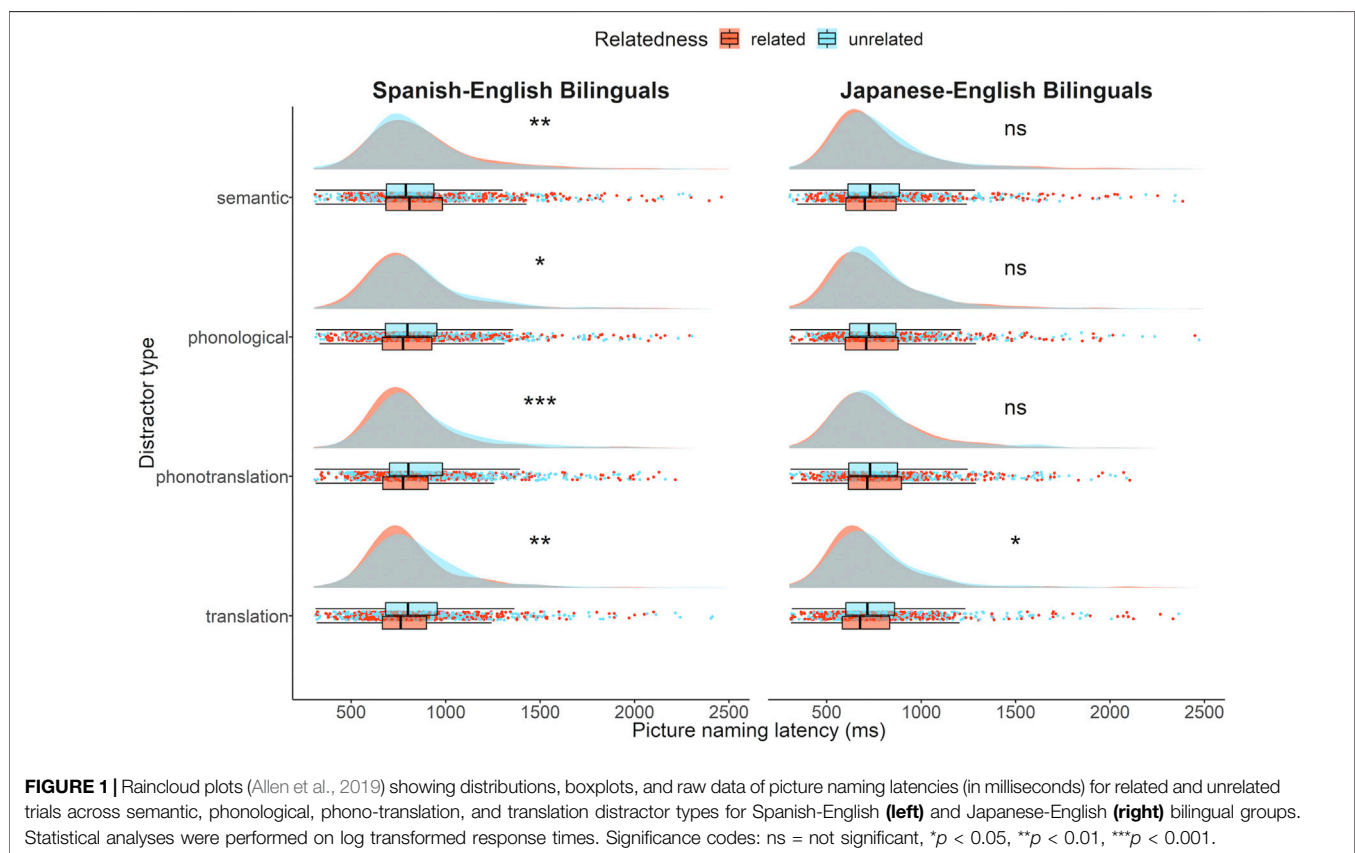
## Results

**Table 3** shows mean picture naming latencies (ms) and accuracy (percent correct) across the four distractor types. Naming latencies associated with related and unrelated conditions are additionally shown in **Figure 1**. Full model outputs are provided (**Supplementary Appendix C**).

<sup>3</sup>Per the editor’s suggestion, we conducted additional analyses including phonological similarity ratings and cognitive measures as predictors, but we found that none of these contributed to model fit. Thus, they were not included in the final models.

**TABLE 3 |** Picture naming mean latencies and accuracy (standard deviation in parenthesis) for related and unrelated trials in Experiment 1 and Experiment 2.

| Distractor type      | Experiment 1               |            | Experiment 2                |            |
|----------------------|----------------------------|------------|-----------------------------|------------|
|                      | Spanish-English bilinguals |            | Japanese-English bilinguals |            |
|                      | Related                    | Unrelated  | Related                     | Unrelated  |
| Latency (ms)         |                            |            |                             |            |
| Semantic             | 881 (305)                  | 844 (275)  | 794 (308)                   | 789 (273)  |
| Phonological         | 834 (285)                  | 861 (281)  | 781 (285)                   | 789 (276)  |
| Phono-translation    | 831 (273)                  | 879 (287)  | 786 (270)                   | 793 (277)  |
| Translation          | 819 (253)                  | 852 (277)  | 748 (284)                   | 772 (269)  |
| Accuracy (% correct) |                            |            |                             |            |
| Semantic             | 93.8 (2.4)                 | 96.0 (2.0) | 94.5 (2.3)                  | 95.2 (2.1) |
| Phonological         | 94.9 (2.2)                 | 95.2 (2.1) | 94.5 (2.3)                  | 94.3 (2.3) |
| Phono-translation    | 95.1 (2.2)                 | 94.9 (2.2) | 95.4 (2.1)                  | 94.8 (2.2) |
| Translation          | 94.7 (2.2)                 | 93.9 (2.4) | 95.2 (2.1)                  | 94.3 (2.3) |



Spanish-English bilinguals named pictures more slowly ( $\beta = -0.02$ ,  $SE = 0.01$ ,  $t = -2.90$ ,  $p = 0.004$ , 95% CI  $[-0.03 \text{ to } -0.01]$ ) and less accurately ( $\beta = 0.67$ ,  $SE = 0.29$ ,  $z = 2.34$ ,  $p = 0.019$ , 95% CI  $[0.11\text{--}1.24]$ ) when distractor words were semantically related than when they were semantically unrelated. Furthermore, relative to unrelated controls, participants named pictures faster when distractor words were phonologically related ( $\beta = 0.02$ ,  $SE = 0.01$ ,  $t = 2.46$ ,  $p = 0.014$ , 95% CI  $[0.00\text{--}0.03]$ ), phonologically related to Spanish translations of English

picture names ( $\beta = 0.02$ ,  $SE = 0.01$ ,  $t = 3.81$ ,  $p < 0.001$ , 95% CI  $[0.01\text{--}0.04]$ ), or Spanish translations of English picture names ( $\beta = 0.02$ ,  $SE = 0.01$ ,  $t = 2.61$ ,  $p = 0.009$ , 95% CI  $[0.00\text{--}0.03]$ ). Picture naming accuracy did not differ between related and unrelated conditions for phonological ( $\beta = 0.08$ ,  $SE = 0.28$ ,  $z = 0.30$ ,  $p = 0.763$ , 95% CI  $[-0.46\text{--}0.63]$ ), phono-translation ( $\beta = -0.04$ ,  $SE = 0.28$ ,  $z = -0.14$ ,  $p = 0.889$ , 95% CI  $[-0.58\text{--}0.51]$ ) or translation ( $\beta = -0.20$ ,  $SE = 0.26$ ,  $z = -0.76$ ,  $p = 0.445$ , 95% CI  $[-0.70\text{--}0.31]$ ) distractor types.

## Discussion

In Experiment 1, Spanish-English bilinguals showed the effect of all the distractor types—phonological facilitation, semantic interference, translation facilitation, and phono-translation facilitation. In other words, we replicated the general pattern of results reported for previous bilingual picture-word interference studies with bilinguals whose two languages share the same Roman alphabets. This replication was found despite the fact that the Spanish-English bilinguals in Experiment 1 did not share the same language profile with respect to age of L2 acquisition, language environment, and language proficiency with the bilinguals tested in the previously published studies (e.g., Hermans et al., 1998; Costa and Caramazza, 1999; Costa et al., 1999; Hermans, 2000; Hermans, 2004; Costa et al., 2003; Knipsky and Amrhein, 2007).

Although we replicated a general pattern of the results of previous bilingual picture-word interference studies for Spanish-English bilinguals, one issue that requires additional discussion concerns the phono-translation effect. The phono-translation distractor words in Experiment 1 for same-script Spanish-English bilinguals produced facilitation rather than interference. Hermans et al. (1998) found an effect of interference for phono-translation distractors for Dutch-English bilinguals. A critical difference between their study and the present study was in the stimulus construction. In Hermans et al., the distractor words were phonologically related, semantically related, phono-translation, or unrelated, whereas the present study included translation distractors in addition to those four types of distractors. When translation names of pictures are included in the task, phono-translation distractors appear to facilitate picture naming rather than interfere the process of speech planning. In other words, just as translation distractor words facilitate the selection of target pictures, phono-translation distractor words activate translation names of pictures and make lexical selection easier. In another study, we did find, like Hermans et al., that Spanish-English bilinguals living in an L2 English environment showed phono-translation interference rather than facilitation when the task did not include translation distractors (Hoshino and Thierry, 2011). This interpretation is also consistent with the results of Hermans et al. (2011) showing that the presence of cross-language activation in a phoneme monitoring task was sensitive to the composition of the experimental materials. Only when there were cognates present in the list context, was cross-language activation observed.

In Experiment 2, we asked whether different-script Japanese-English bilinguals would also show the same pattern of the results as same-script Spanish-English bilinguals in Experiment 1. In the absence of a language cue, Spanish-English and Japanese-English bilinguals perform similarly on a simple picture naming task (Hoshino and Kroll, 2008). If Japanese distractors words provide a language cue to the language of production, then Japanese-English bilinguals should show smaller cross-language effects than Spanish-English bilinguals.

## EXPERIMENT 2: JAPANESE-ENGLISH BILINGUALS

### Method

#### Participants

Thirty-nine Japanese-English bilinguals who were living in the L2 environment at the time of testing participated in Experiment 2. They completed the same set of tasks as Spanish-English bilinguals in Experiment 1. The characteristics of the Japanese-English bilinguals are summarized in **Table 1**.

### Materials

#### Pictures

The pictures were identical to those used in Experiment 1.

#### Distractor Words

For each of the pictures, four types of distractor words were selected in Japanese: phonologically related to the English picture name, semantically related to the English picture name, Japanese translation name of the English picture name, phonologically related to the Japanese translation name of the English picture name (phono-translation) (see **Table 4** for examples). Similar to Experiment 1, the following criteria were used to select each type of distractor: 1) the distractors that were phonologically related to the English picture name or to the Japanese translation of the English picture name were matched on phonological onset with the English picture name and were not also semantically related to the target picture; 2) words were typically written in kanji; 4) the semantically related distractors were identical to those in Experiment 1 (i.e., the English translation of the semantically related distractors were the same). Each of these related distractors had an unrelated control that was matched item-by-item based on length (number of characters and syllables) and frequency [all  $p$ s > 0.10] (see **Table 4**). Each of the filler pictures also had a distractor word in Japanese. The complete set of the experimental items is provided (**Supplementary Appendix B**). The organization of the lists and blocks was identical to the one in Experiment 1.

#### Norming

Although phonologically related and phono-translation distractors were matched on phonological onset with English picture names in both Experiment 1 and Experiment 2, an additional measure of phonological similarity of the distractors and picture names was obtained to ensure that observed cross-language differences between Experiment 1 and Experiment 2, if any, were due to script but not to differences in phonological similarity. Fifteen English monolinguals who had not studied Spanish and 16 English monolinguals who had not studied Japanese were asked to rate sound pairs according to how similar two words sounded on a 7-point Likert scale with “1” being completely different and “7” being identical.<sup>4</sup> The sound pairs consisted of an English picture name and its phonologically

<sup>4</sup>Two independent groups of English monolinguals were recruited to minimize the effect of speakers.

**TABLE 4 |** Examples of distractors for the picture “envelope” by distractor type and relatedness and characteristics of Japanese distractors.

| Distractor type   | Examples                   |                              | Frequency |           | Length    |           |
|-------------------|----------------------------|------------------------------|-----------|-----------|-----------|-----------|
|                   | Related                    | Unrelated                    | Related   | Unrelated | Related   | Unrelated |
| Phonological      | 煙突 (chimney)<br>/eNtotu/   | 大根 (radish)<br>/daikoN/      | 3.375     | 3.390     | 1.7 (2.7) | 1.7 (2.7) |
| Semantic          | 葉書 (postcard)<br>/hagaki/  | 毛虫 (caterpillar)<br>/kemusi/ | 3.409     | 3.390     | 1.5 (2.7) | 1.5 (2.7) |
| Translation       | 封筒 (envelope)<br>/huRtoR/  | 花火 (firework)<br>/hanabi/    | 3.642     | 3.390     | 1.3 (2.8) | 1.3 (2.7) |
| Phono-translation | 風鈴 (wind-bell)<br>/huRriN/ | 王冠 (crown)<br>/oRkaN/        | 3.670     | 3.390     | 1.7 (2.7) | 1.7 (2.7) |

(1) The translation of the distractor word and its phonemic transcription are given in parentheses and in slashes, respectively. (2) The number of characters is provided without parentheses and the number of morae is provided with parentheses. Frequency was from Amano and Kondo (2000) and log frequency was computed based on the values from the source.

**TABLE 5 |** Similarity ratings by English monolinguals as a function of distractor type and relatedness.

| Distractor type   | Spanish   |           | Japanese  |           |
|-------------------|-----------|-----------|-----------|-----------|
|                   | Related   | Unrelated | Related   | Unrelated |
| Phonological      | 3.6 (1.3) | 1.2 (0.2) | 4.6 (1.3) | 1.3 (0.2) |
| Phono-translation | 4.1 (0.5) | 1.4 (0.5) | 4.3 (0.7) | 1.4 (0.3) |

Standard deviations are in parentheses.

related or unrelated Spanish/Japanese distractor word and pairs consisting of a Spanish/Japanese picture name and its phonologically related or unrelated Spanish/Japanese distractor word. The English picture names were recorded by a female native speaker of English and the Spanish and Japanese distractor words were recorded by female native speakers of Spanish and Japanese, respectively. The sound file of each distractor word in Spanish or in Japanese was combined with that of the target English picture name or the target Spanish/Japanese picture name. A set of English-Spanish/Japanese sound pairs and a set of Spanish-Spanish or Japanese-Japanese sound pairs were created. Each stimulus set consisted of 32 sound pairs and therefore, each participant received 64 sound pairs.

The mean ratings for each condition are summarized by sound pairs in **Table 5**. A critical result in the norming experiment was that monolingual English speakers perceived the phonological similarity of related pairs to be greater than unrelated pairs in the phonological condition [ $t(15) = 7.61, p < 0.001$  for English-Spanish pairs;  $t(15) = 11.25, p < 0.001$  for English-Japanese pairs] and in the phono-translation condition [ $t(15) = 17.68, p < 0.001$  for Spanish-Spanish pairs;  $t(15) = 14.31, p < 0.001$  for Japanese-Japanese pairs], regardless of language pairs. Although care was taken to ensure that phonological similarity would be similar across experiments, the English names of pictures and their phonologically related Japanese distractors were rated as more similar than those of the English names of pictures and their phonologically related Spanish distractors [ $t(15) = 3.12, p < 0.01$ ]. However, it is important to note that if this difference in phonological similarity of items influences bilingual performance, then Japanese-English bilinguals in Experiment 2 should show greater phonological facilitation than Spanish-English bilinguals

in Experiment 1, which would counter the predicted reduction of distractor effects for different-script bilinguals.

## Procedure

The same procedure was used as in Experiment 1 except for that the distractor words were Japanese, not Spanish.

## Data Trimming Procedure

The data trimming procedure was identical to Experiment 1. Errors (4.8%), technical errors (<0.1%), and outliers (1.0%) were excluded from the data analyses.

## Data Analysis

The same procedure was used as in Experiment 1.

## Results

**Table 3** shows mean picture naming latencies (ms) and accuracy (percent correct) across the four distractor types. Naming latencies associated with related and unrelated conditions are additionally shown in **Figure 1**. Full model outputs are provided (**Supplementary Appendix C**).

In contrast to Spanish-English bilinguals, Japanese-English bilinguals did not show a semantic interference effect (naming latency:  $\beta = 0.00$ ,  $SE = 0.01$ ,  $t = 0.29$ ,  $p = 0.772$ , 95% CI  $[-0.01-0.02]$ ; naming accuracy:  $\beta = 0.21$ ,  $SE = 0.33$ ,  $z = 0.65$ ,  $p = 0.516$ , 95% CI  $[-0.43-0.85]$ ). They also did not show relatedness effects in naming latency or accuracy for phonological (naming latency:  $\beta = 0.01$ ,  $SE = 0.01$ ,  $t = 1.21$ ,  $p = 0.228$ , 95% CI  $[-0.01-0.02]$ ; naming accuracy:  $\beta = -0.07$ ,  $SE = 0.30$ ,  $z = -0.22$ ,  $p = 0.826$ , 95% CI  $[-0.65-0.52]$ ) or phono-translation (naming latency:  $\beta = 0.00$ ,  $SE = 0.01$ ,  $t = 0.26$ ,  $p = 0.796$ , 95% CI  $[-0.01-0.02]$ ; naming accuracy:  $\beta = -0.16$ ,  $SE = 0.32$ ,  $z = -0.48$ ,  $p = 0.628$ , 95% CI  $[-0.79-0.48]$ ) distractor types. However, Japanese-English bilinguals named pictures faster when distractor words were Japanese translations of English picture names than when they were unrelated ( $\beta = 0.02$ ,  $SE = 0.01$ ,  $t = 2.29$ ,  $p = 0.022$ , 95% CI  $[0.00-0.03]$ ), but naming accuracy did not differ in this condition ( $\beta = -0.25$ ,  $SE = 0.32$ ,  $z = -0.79$ ,  $p = 0.431$ , 95% CI  $[-0.87-0.37]$ ).

## Discussion

Unlike findings for simple picture naming, in which the performance of Spanish-English and Japanese-English

bilinguals was identical (Hoshino and Kroll, 2008), Spanish-English bilinguals (Experiment 1) and Japanese-English bilinguals (Experiment 2) performed differently on picture naming in the presence of language-specific distractor words. Similar to Spanish-English bilinguals in Experiment 1, Japanese-English bilinguals showed translation facilitation. Unlike Spanish-English bilinguals, however, they did not show phonological facilitation, semantic interference, and phono-translation facilitation. The absence of phonological and phono-translation effects might be due to the characteristics of kanji scripts. Past research suggests that phonology is specified earlier in kana than in kanji, whereas semantic access occurs earlier in kanji than in kana (e.g., Yamada, 1998; Ischebeck, 2004; Chen et al., 2007). A critical finding in the present study is that Japanese-English bilinguals did not show semantic interference even with kanji distractor words, and this finding is consistent with the results of the picture-word interference studies with bimodal bilinguals (Giezen and Emmorey, 2016; Emmorey et al., 2020b). In sum, these results suggest that when the distinctive script is present in the task, different-script bilinguals are able to exploit the perceptual information as a cue to allow language selection to occur earlier in speech planning relative to same-script bilinguals.

## GENERAL DISCUSSION

The goal of the present study was to determine whether the degree of cross-language activation and the locus of language selection could be modulated by script when the task included an overt written lexical form. Spanish-English bilinguals (Experiment 1) and Japanese-English bilinguals (Experiment 2) named pictures in their L2 English while ignoring visually presented L1 (Spanish/Japanese) distractor words. The distractor words were manipulated in relation to the picture to create four conditions: phonological, semantic, translation, and phono-translation. Unlike findings for simple picture naming, Spanish-English and Japanese-English bilinguals in the present study performed differently on picture naming in the presence of language-specific distractor words. In the picture-word interference task, both groups showed translation facilitation, whereas only Spanish-English bilinguals demonstrated phonological facilitation, semantic interference, and phono-translation facilitation. In other words, we replicated a general pattern of the results of previous bilingual picture-word interference studies for same-script Spanish-English bilinguals, whereas the pattern of results differed for different-script Japanese-English bilinguals.

We now consider why Japanese-English bilinguals showed translation facilitation, but not phonological facilitation, semantic interference, and phono-translation facilitation. We argue that when the distinctive script is present in the task, different-script bilinguals are able to exploit the perceptual information as a cue to allow language selection to occur earlier in speech planning relative to same-script bilinguals. According to this account, lexical candidates from both languages are activated for a very brief period of time but speech planning then becomes language-selective such that only lexical candidates from the target

language (i.e., English) compete for selection. As can be seen in **Table 3**, Japanese-English bilinguals were faster to name pictures than Spanish-English bilinguals although the two groups were matched on verbal fluency, which was a measure of productive skills, and if anything, Spanish-English bilinguals appeared more proficient in English on other measures. This difference might also reflect the early language selection by Japanese-English bilinguals.<sup>5</sup> Indeed, this account is in line with studies showing that Chinese-English bilinguals named images culturally matched with the language to be spoken faster than those culturally mismatched (e.g., Jared et al., 2013; Li et al., 2013), that bimodal bilinguals did not show semantic interference in naming pictures in American Sign Language (ASL) while ignoring English distractor words (Giezen and Emmorey, 2016; Emmorey et al., 2020a), and that Hebrew-English bilinguals read aloud mixed-language texts more accurately than Spanish-English bilinguals (Fadlon et al., 2019). This is also compatible with studies showing that different-script bilinguals coactivate the non-target language in a semantic relatedness judgment task where the non-target language is only implicitly available (Thierry and Wu, 2007; Degani et al., 2018).

On this account of “early selection”, distractors are unlikely to have an effect except when they are the translation of the picture name because the phonological and semantic representation of the distractor will be available only after language selection has occurred. Why would there be an effect for translation equivalents but not for other semantic related distractors? It appears that the semantic activation of the picture itself primed the recognition of the distractor when it was the translation, i.e., the name of the picture, to create convergence among related conceptual nodes. There are two results in the past literature that suggest that resonance among activated lexical codes may be a critical factor. Previous studies of bilingual word recognition (e.g., Dijkstra et al., 1998) have shown that cognates produce more robust cross-language effects than interlingual homographs. The analogy with the present result is that semantic conflicts across languages can sometimes be ignored whereas semantic convergence can almost never be ignored (see Schwartz and Kroll, 2006, for an illustration of the same phenomenon within sentence context). Another feature of the translation facilitation effect that is relevant to the present discussion is that in experiments in which the SOA has been manipulated between the presentation of the picture and distractor, translation facilitation only occurs very early. Because the data for Japanese-English bilinguals suggest that they are not able to selectively ignore the Japanese distractor

<sup>5</sup>Alternatively, the absence of phonological, semantic, and phono-translation effects for Japanese-English bilinguals might have been due to individual differences in the ability to ignore irrelevant information and in the availability of the amount of processing resources. As shown in **Table 1**, the Japanese-English bilinguals produced a smaller Simon effect and a larger operation span than the Spanish-English bilinguals. These cognitive measures did not improve model fit in both Experiment 1 and Experiment 2. However, it is critical to investigate the extent to which cognitive control abilities as well as script differences can contribute to cross-language distractor effects in future research.

words until some processing of the distractor has occurred, those processes that reflect early interactions between the bottom-up activation of the word and the top-down information engaged by the picture, are likely to survive the cross-language script difference. In Emmorey et al. (2020a), in fact, ASL-English bilinguals showed the effect of relatedness in the time window of 200–300 ms for the translation condition but not for the semantic condition.

In the present study, both the Spanish-English and the Japanese-English bilinguals produced translation facilitation, whereas only the Spanish-English bilinguals produced phonological facilitation, semantic interference, and phono-translation facilitation. The pattern of these results suggests that script differences modulate cross-language activation during production when the written lexical form is perceptually available in one language. The distinctive script appears to serve as a language cue to direct attention to the lemmas in the target language alone at an earlier stage of speech planning, a finding that is in line with the assumption that the flow of activation is nonselective but the manner of language selection may be language-specific (e.g., Costa and Caramazza, 1999; Costa et al., 1999). In other words, even if there is cross-language activation, the activation of lexical candidates from the non-response language does not necessarily interfere with lexical selection in the intended language if there is a basis on which the language of speaking can be selected in advance. However, the fact that the presence of the script difference alone was not sufficient to create an entirely selective, monolingual-like situation for the Japanese-English bilinguals, is compatible with a model that assumes that all activated lexical candidates from both languages compete for selection (e.g., Green, 1998). Different scripts may function to inhibit unintended alternatives earlier in the process, thereby eliminating phonological facilitation, semantic interference, and phono-translation facilitation, but they cannot override cross-language activation entirely.

Alternatively, the Response Exclusion Hypothesis assumes that competition occurs at a post-lexical level, not at a lexical level (Mahon et al., 2007). On this account, distractor words activate their representations in the articulators prior to the picture. Because the articulators are a single-channel buffer, non-target representations need to be excluded to articulate the target picture name. When more features are shared between the target picture name and distractors, it takes longer to reject non-target candidates from the buffer. The absence of the distractor effects other than the translation facilitation for Japanese-English bilinguals may be due to the fact that it takes longer to have access to the phonological properties of kanji distractor words. Unlike alphabetic writing systems, the pronunciation of a kanji character is not transparent because its components do not correspond to the individual phonemes of the pronunciation, which makes it take longer to retrieve the phonology of the character. By the time kanji distractors activate the representations in the articulators, the language of production is already selected.

Another possible account is based on a model that assumes that lexical alternatives in the non-target language are further from the selection criteria (threshold) and thus are rejected more

easily than alternatives in the response language (Finkbeiner et al., 2006). This threshold model posits that the bilingual's intention to speak in one language activates the target language more strongly than the non-target language and lexical candidates in the target language will reach the threshold for selection more quickly. The absence of phonological, semantic, and phono-translation effects for Japanese-English bilinguals might be explained by the threshold account if we assume that the distinctive script does not meet the selection criterion and lexical alternatives in the nonresponse language can be rejected rapidly. However, if the absence of phonological, semantic, and phono-translation effects were due to the adjustment of selection criteria, then translation facilitation should also have not been obtained. It is important to note that in the present study, the distractor conditions were mixed so that the Japanese-English bilinguals could not simply set a different threshold strategically, depending upon the type of distractor words.

In summary, the present study replicated a general pattern of the results of past bilingual picture-word interference studies for same-script bilinguals (Spanish-English) but only partly for different-script bilinguals (Japanese-English). This specific pattern of the present results suggests that when script is perceptually available, the degree of cross-language activation and the locus of language selection is modulated by script differences between the bilingual's two languages. Based on the obtained results, we have argued that the flow of activation in the mechanism of language production is fundamentally nonselective. Language-specific differences such as script can serve as a language cue to allow the bilingual to select the intended language earlier in the process of speech planning when they are perceptually available. That is, these findings suggest that the locus of language selection in bilingual speech planning is “not” fixed (see Kroll et al., 2006 for a review). The fact that some, but not all, distractor conditions were effective for the Japanese-English bilinguals is consistent with an account of bilingual production in which activated candidates in the non-target languages are suppressed earlier in speech planning when language status is available. If different-script bilinguals had been better able to attend to the target language from the start, then no effects of the distractors should have been observed. In future research, it will be critical to further examine the time course of language/lexical selection as a function of type of bilingualism (same script vs. different script).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Pennsylvania State University Institutional Review Board. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

This study was part of the dissertation project conducted by NH under the supervision of JK. AB-M and CN-T analyzed the data for the manuscript. All the authors contributed to the writing of the manuscript.

## FUNDING

This research was supported in part by NSF Dissertation Grant BCS-0518814 to NH and JK. The writing of this article was supported in part by JSPS KAKENHI Grant Numbers JP24730624 and JP20K00756 to NH, and by NIH grant F32-AG064810 to AB-M, NSF Grant BCS-1824072 and NIH Grant

F31HD098783 to CN-T, and by NSF Grant OISE-1545900 to JK.

## ACKNOWLEDGMENTS

We thank Kristin Dresner, Kay Holtzinger, and Hiroko Kitajima for research assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.668381/full#supplementary-material>

## REFERENCES

- Alameda, J. R., and Cuetos, F. (1995). *Diccionario de frecuencias de las unidas de castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.
- Alario, F.-X., Perre, L., Castel, C., and Ziegler, J. C. (2007). The Role of Orthography in Speech Production Revisited. *Cognition* 102, 464–475. doi:10.1016/j.cognition.2006.02.002
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., and Kievit, R. A. (2019). Raincloud Plots: a Multi-Platform Tool for Robust Data Visualization. *Wellcome Open Res.* 4, 63. doi:10.12688/wellcomeopenres.15191.1
- Amano, S., and Kondo, T. (2000). *Nihongo No Goi Tokusei [Lexical Properties of Japanese]*. Tokyo: Sanseido.
- Azuma, T., and Van Orden, G. C. (1997). Why Safe Is Better Than Fast: The Relatedness of a Word's Meanings Affects Lexical Decision Times. *J. Mem. Lang.* 36, 484–504. doi:10.1006/jmla.1997.2502
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects Modeling with Crossed Random Effects for Subjects and Items. *J. Mem. Lang.* 59, 390–412. doi:10.1016/j.jml.2007.12.005
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *J. Mem. Lang.* 68, 1–43. doi:10.1016/j.jml.2012.11.001
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48. doi:10.18637/jss.v067.i01
- Bi, Y., Xu, Y., and Caramazza, A. (2009). Orthographic and Phonological Effects in the Picture-word Interference Paradigm: Evidence from a Logographic Language. *Appl. Psycholinguistics* 30, 637–658. doi:10.1017/s0142716409990051
- Bialystok, E., Craik, F. I. M., Klein, R., and Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence from the Simon Task. *Psychol. Aging* 19, 290–303. doi:10.1037/0882-7974.19.2.290
- Blanco-Elorrieta, E., Emmorey, K., and Pykkänen, L. (2018). Language Switching Decomposed through MEG and Evidence from Bimodal Bilinguals. *Proc. Natl. Acad. Sci. USA* 115, 9708–9713. doi:10.1073/pnas.1809779115
- Boukadi, M., Davies, R. A. I., and Wilson, M. A. (2015). Bilingual Lexical Selection as a Dynamic Process: Evidence from Arabic-French Bilinguals. *Can. J. Exp. Psychology/Revue canadienne de Psychol. expérimentale* 69, 297–313. doi:10.1037/cep0000063
- Brauer, M. (1998). “Stroop Interference in Bilinguals: The Role of Similarity between the Two Languages,” in *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*. Editors A. F. Healy and L. E. Bourne (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 317–337.
- Chen, H.-c., and Ho, C. (1986). Development of Stroop Interference in Chinese-English Bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 12, 397–401. doi:10.1037/0278-7393.12.3.397
- Chen, H.-C., Yamauchi, T., Tamaoka, K., and Vaid, J. (2007). Homophonic and Semantic Priming of Japanese Kanji Words: A Time Course Study. *Psychon. Bull. Rev.* 14, 64–69. doi:10.3758/bf03194029
- Chéreau, C., Gaskell, M. G., and Dumay, N. (2007). Reading Spoken Words: Orthographic Effects in Auditory Priming. *Cognition* 102, 341–360. doi:10.1016/j.cognition.2006.01.001
- Costa, A., and Caramazza, A. (1999). Is Lexical Selection in Bilingual Speech Production Language-specific? Further Evidence from Spanish-English and English-Spanish Bilinguals. *Bilingualism* 2, 231–244. doi:10.1017/s1366728999000334
- Costa, A., Colomé, À., Gómez, O., and Sebastián-gallés, N. (2003). Another Look at Cross-Language Competition in Bilingual Speech Production: Lexical and Phonological Factors. *Bilingualism* 6, 167–179. doi:10.1017/s1366728903001111
- Costa, A. (2005). “Lexical Access in Bilingual Production,” in *Handbook of Bilingualism: Psycholinguistic Approaches*. Editors J. F. Kroll and A. M. B. De Groot (New York: Oxford University Press), 308–325.
- Costa, A., Miozzo, M., and Caramazza, A. (1999). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection?. *J. Mem. Lang.* 41, 365–397. doi:10.1006/jmla.1999.2651
- Damian, M. F., and Bowers, J. S. (2009). Assessing the Role of Orthography in Speech Perception and Production: Evidence from Picture-word Interference Tasks. *Eur. J. Cogn. Psychol.* 21, 581–598. doi:10.1080/09541440801896007
- Damian, M. F., and Bowers, J. S. (2003). Effects of Orthography on Speech Production in a Form-Preparation Paradigm. *J. Mem. Lang.* 49, 119–132. doi:10.1016/s0749-596x(03)00008-1
- Degani, T., Prior, A., and Hajjira, W. (2018). Cross-language Semantic Influences in Different Script Bilinguals. *Bilingualism* 21, 782–804. doi:10.1017/s1366728917000311
- Dijkstra, T., Van Jaarsveld, H., and Brinke, S. T. (1998). Interlingual Homograph Recognition: Effects of Task Demands and Language Intermixing. *Bilingualism* 1, 51–66. doi:10.1017/s1366728998000121
- Emmorey, K., Li, C., Petrich, J., and Gollan, T. H. (2020a). Turning Languages on and off: Switching into and Out of Code-Blends Reveals the Nature of Bilingual Language Control. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 443–454. doi:10.1037/xlm0000734
- Emmorey, K., Mott, M., Meade, G., Holcomb, P. J., and Midgley, K. J. (2020b). Lexical Selection in Bimodal Bilinguals: ERP Evidence from Picture-word Interference. *Lang. Cogn. Neurosci.* 1, 1–15. doi:10.1080/23273798.2020.1821905
- Fadlon, J., Li, C., Prior, A., and Gollan, T. H. (2019). Using What's There: Bilinguals Adaptively Rely on Orthographic and Color Cues to Achieve Language Control. *Cognition* 191, 103990. doi:10.1016/j.cognition.2019.06.002
- Finkbeiner, M., Gollan, T. H., and Caramazza, A. (2006). Lexical Access in Bilingual Speakers: What's the (Hard) Problem? *Bilingualism* 9, 153–166. doi:10.1017/s1366728906002501
- Giezen, M. R., and Emmorey, K. (2016). Language Co-activation and Lexical Selection in Bimodal Bilinguals: Evidence from Picture-word Interference. *Bilingualism* 19, 264–276. doi:10.1017/s1366728915000097
- Green, D. W. (1998). Mental Control of the Bilingual Lexico-Semantic System. *Bilingualism* 1, 67–81. doi:10.1017/s1366728998000133

- Guo, T., Liu, H., Misra, M., and Kroll, J. F. (2011). Local and Global Inhibition in Bilingual Word Production: fMRI Evidence from Chinese-English Bilinguals. *NeuroImage* 56 (4), 2300–2309. doi:10.1016/j.neuroimage.2011.03.049
- Hermans, D. (2004). Between-language Identity Effects in Picture-word Interference Tasks: A challenge for Language-Nonspecific or Language-specific Models of Lexical Access? *Int. J. Bilingualism* 8, 115–125. doi:10.1177/13670069040080020101
- Hermans, D., Bongaerts, T., De Bot, K., and Schreuder, R. (1998). Producing Words in a Foreign Language: Can Speakers Prevent Interference from Their First Language? *Bilingualism* 1, 213–229. doi:10.1017/s1366728998000364
- Hermans, D., Ormel, E., van Besselaar, R., and Van Hell, J. (2011). Lexical Activation in Bilinguals' Speech Production Is Dynamic: How Language Ambiguous Words Can Affect Cross-Language Activation. *Lang. Cogn. Process.* 26, 1687–1709. doi:10.1080/01690965.2010.530411
- Hermans, D. (2000). *Word Production in a Foreign Language*. Nijmegen, Netherlands: Unpublished doctoral dissertation, University of Nijmegen.
- Hoshino, N., and Kroll, J. F. (2008). Cognate Effects in Picture Naming: Does Cross-Language Activation Survive a Change of Script? *Cognition* 106, 501–511. doi:10.1016/j.cognition.2007.02.001
- Hoshino, N., and Thierry, G. (2011). Language Selection in Bilingual Word Production: Electrophysiological Evidence for Cross-Language Competition. *Brain Res.* 1371, 100–109. doi:10.1016/j.brainres.2010.11.053
- Ischebeck, A. K. (2004). *Processing Phonographic and Morphographic Script: Similarities and Differences*. Nijmegen, Netherlands: Unpublished doctoral dissertation, University of Nijmegen.
- Jared, D., Pei Yun Poh, R., and Paivio, A. (2013). L1 and L2 Picture Naming in Mandarin-English Bilinguals: A Test of Bilingual Dual Coding Theory. *Bilingualism* 16, 383–396. doi:10.1017/s1366728912000685
- Knopsky, A. C., and Amrhein, P. C. (2007). Phonological Facilitation through Translation in a Bilingual Picture-Naming Task. *Bilingualism* 10, 211–223. doi:10.1017/s1366728907003033
- Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language Selection in Bilingual Speech: Evidence for Inhibitory Processes. *Acta Psychologica* 128, 416–430. doi:10.1016/j.actpsy.2008.02.001
- Kroll, J. F., Bobb, S. C., and Wodniecka, Z. (2006). Language Selectivity Is the Exception, Not the Rule: Arguments against a Fixed Locus of Language Selection in Bilingual Speech. *Bilingualism* 9, 119–135. doi:10.1017/s1366728906002483
- La Heij, W., De Bruyn, E., Elens, E., Hartsuiker, R., Helaha, D., and Van Schelven, L. (1990). Orthographic Facilitation and Categorical Interference in a Word-Translation Variant of the Stroop Task. *Can. J. Psychology/Revue canadienne de Psychol.* 44, 76–83. doi:10.1037/h0084236
- Lee, T. M. C., and Chan, C. C. H. (2000). Stroop Interference in Chinese and English. *J. Clin. Exp. Neuropsychol.* 22, 465–471. doi:10.1076/1380-3395(200008)22:4;1-0;ft465
- Li, Y., Yang, J., Suzanne Scherf, K., and Li, P. (2013). Two Faces, Two Languages: An fMRI Study of Bilingual Picture Naming. *Brain Lang.* 127, 452–462. doi:10.1016/j.bandl.2013.09.005
- Linck, J. A., Hoshino, N., and Kroll, J. F. (2008). Cross-language Lexical Processes and Inhibitory Control. *Ment. Lex* 3, 349–374. doi:10.1075/ml.3.3.06lin
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing Access to the Native Language while Immersed in a Second Language: Evidence for the Role of Inhibition in Second-Language Learning. *Psychol. Sci.* 20, 1507–1515. doi:10.1111/j.1467-9280.2009.02480.x
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., and Caramazza, A. (2007). Lexical Selection Is Not by Competition: a Reinterpretation of Semantic Interference and Facilitation Effects in the Picture-word Interference Paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 33 (3), 503–535. doi:10.1037/0278-7393.33.3.503
- Meuter, R. F. I., and Allport, A. (1999). Bilingual Language Switching in Naming: Asymmetrical Costs of Language Selection. *J. Mem. Lang.* 40, 25–40. doi:10.1006/jmla.1998.2602
- Miller, N. A., and Kroll, J. F. (2002). Stroop Effects in Bilingual Translation. *Mem. Cogn.* 30, 614–628. doi:10.3758/bf03194963
- Misra, M., Guo, T., Bobb, S. C., and Kroll, J. F. (2012). When Bilinguals Choose a Single Word to Speak: Electrophysiological Evidence for Inhibition of the Native Language. *J. Mem. Lang.* 67, 224–237. doi:10.1016/j.jml.2012.05.001
- Moon, J., and Jiang, N. (2012). Non-selective Lexical Access in Different-Script Bilinguals. *Bilingualism* 15, 173–180. doi:10.1017/s1366728911000022
- Pattamadilok, C., Perre, L., Dufau, S., and Ziegler, J. C. (2009). On-line Orthographic Influences on Spoken Language in a Semantic Task. *J. Cogn. Neurosci.* 21, 169–179. doi:10.1162/jocn.2009.21014
- Philipp, A. M., and Koch, I. (2009). Inhibition in Language Switching: What Is Inhibited when Switching between Languages in Naming Tasks? *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1187–1195. doi:10.1037/a0016376
- Rastle, K., McCormick, S. F., Bayliss, L., and Davis, C. J. (2011). Orthography Influences the Perception and Production of Speech. *J. Exp. Psychol. Learn. Mem. Cogn.* 37 (6), 1588–1594. doi:10.1037/a0024833
- Roelofs, A. (2006). The Influence of Spelling on Phonological Encoding in Word reading, Object Naming, and Word Generation. *Psychon. Bull. Rev.* 13, 33–37. doi:10.3758/bf03193809
- Schwartz, A. I., and Kroll, J. F. (2006). Bilingual Lexical Activation in Sentence Context. *J. Mem. Lang.* 55, 197–212. doi:10.1016/j.jml.2006.03.004
- Smith, M. C., and Kirsner, K. (1982). Language and Orthography as Irrelevant Features in Colour-word and Picture-word Stroop Interference. *The Q. J. Exp. Psychol. Section A* 34, 153–170. doi:10.1080/14640748208400864
- Snodgrass, J. G., and Vanderwart, M. (1980). A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215. doi:10.1037/0278-7393.6.2.174
- Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., et al. (2003). Timed Picture Naming: Extended Norms and Validation against Previous Studies. *Behav. Res. Methods Instr. Comput.* 35, 621–633. doi:10.3758/bf03195542
- Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A New On-Line Resource for Psycholinguistic Studies. *J. Mem. Lang.* 51, 247–250. doi:10.1016/j.jml.2004.03.002
- Tan, L. H., and Perfetti, C. A. (1999). Phonological Activation in Visual Identification of Chinese Two-Character Words. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 382–393. doi:10.1037/0278-7393.25.2.382
- Thierry, G., and Wu, Y. J. (2007). Brain Potentials Reveal Unconscious Translation during Foreign-Language Comprehension. *Proc. Natl. Acad. Sci.* 104, 12530–12535. doi:10.1073/pnas.0609927104
- Tokowicz, N., Michael, E. B., and Kroll, J. F. (2004). The Roles of Study-Abroad Experience and Working-Memory Capacity in the Types of Errors Made during Translation. *Bilingualism* 7, 255–272. doi:10.1017/s1366728904001634
- Van Heuven, W. J., Conklin, K., Coderre, E. L., Guo, T., and Dijkstra, T. (2011). The Influence of Cross-Language Similarity on Within-And Between-Language Stroop Effects in Trilinguals. *Front. Psychol.* 2, 374. doi:10.3389/fpsyg.2011.00374
- Yamada, J. (1998). The Time Course of Semantic and Phonological Access in Naming Kanji and Kana Words. *Reading and Writing* 10, 425–437. doi:10.1023/a:1008095920749
- Zhang, Q., Chen, H. C., Weekes, B. S., and Yang, Y. (2009). Independent Effects of Orthographic and Phonological Facilitation on Spoken Word Production in Mandarin. *Lang. Speech* 52 (1), 113–126. doi:10.1177/0023830908099885
- Zhang, Q., and Weekes, B. S. (2009). Orthographic Facilitation Effects on Spoken Word Production: Evidence from Chinese. *Lang. Cogn. Process.* 24, 1082–1096. doi:10.1080/01690960802042133
- Ziegler, J. C., Muneaux, M., and Grainger, J. (2003). Neighborhood Effects in Auditory Word Recognition: Phonological Competition and Orthographic Facilitation. *J. Mem. Lang.* 48, 779–793. doi:10.1016/s0749-596x(03)00006-8

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hoshino, Beatty-Martínez, Navarro-Torres and Kroll. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Cross-Language Influences in the Processing of Multiword Expressions: From a First Language to Second and Back

Lingli Du<sup>1,2\*</sup>, Irina Elgort<sup>2</sup> and Anna Siyanova-Chanturia<sup>2,3</sup>

<sup>1</sup> School of Foreign Languages, Henan University of Technology, Zhengzhou, China, <sup>2</sup> School of Linguistics and Applied Language Studies, Te Herenga Waka – Victoria University of Wellington, Wellington, New Zealand, <sup>3</sup> College of Foreign Languages, Ocean University of China, Qingdao, China

## OPEN ACCESS

### Edited by:

Tamar Degani,  
University of Haifa, Israel

### Reviewed by:

Manuel F. Pulido,  
Pennsylvania State University,  
United States  
Junko Yamashita,  
Nagoya University, Japan

### \*Correspondence:

Lingli Du  
lingli.du@vuw.ac.nz

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 February 2021

**Accepted:** 06 May 2021

**Published:** 24 June 2021

### Citation:

Du L, Elgort I and  
Siyanova-Chanturia A (2021)  
Cross-Language Influences  
in the Processing of Multiword  
Expressions: From a First Language  
to Second and Back.  
Front. Psychol. 12:666520.  
doi: 10.3389/fpsyg.2021.666520

The present study investigated cross-language influences in the processing of binomial expressions (*knife and fork*), from a first language (L1) to a second language (L2) and from L2 to L1. Two groups of unbalanced bilinguals (Chinese/L1-English/L2 and English/L1-Chinese/L2) and a control group of English monolinguals performed a visual lexical decision task that incorporated unmasked priming. To assess cross-language influences, we used three types of expressions: congruent binomials (English binomials that have translation equivalents in Chinese), English-only binomials, and Chinese-only binomials translated into English. Lexical decision latencies to the last word (*fork*) in a binomial (*knife and fork*) were compared with response latencies to the same word in a matched control phrase (*spoon and fork*). We found that (1) Chinese-English bilinguals showed a significant priming effect for congruent binomials but no facilitation for English-only binomials, (2) English-Chinese bilinguals showed a trend toward priming for congruent binomials, which did not reach statistical significance, and no priming for English-only binomials, (3) English monolinguals showed comparable priming for congruent and English-only binomials. With respect to the Chinese-only binomials, none of the three participant groups showed priming for translated Chinese-only binomials over controls. These findings suggest that L1 influences the processing of L2 binomials, and that there may be some cross-linguistic influence in the opposite direction, i.e., from L2 to L1, although to a lesser extent.

**Keywords:** multiword expressions, binomials, cross-language influence, congruency, frequency, English, Chinese, priming

## INTRODUCTION

Research in bilingual language processing extends beyond single words, to lexical units known as multiword expressions (MWEs), such as idioms (*kick the bucket*) and collocations (*strong tea*). Bilingual research shows that a bilingual's first language (L1) can influence the processing of a second language (e.g., Keatley et al., 1994; Kim and Davis, 2003; Duyck, 2005; Schoonbaert et al., 2007), and that a non-dominant second language (L2) can also influence the processing in the dominant L1 (Jiang, 1999; van Hell and Dijkstra, 2002; Schoonbaert et al., 2009). Specifically, bilingual processing has been found to entail cross-language activation (e.g., Kroll and Bialystok, 2013; Conklin, 2020; Whitford and Titone, 2019). In bilingual processing

beyond the word level, L1 has been found to influence the processing of MWEs in an L2 (Wolter and Gyllstad, 2011, 2013; Carrol and Conklin, 2014, 2017; Carrol et al., 2016), but cross-linguistic influences in the L2-L1 direction are less clear. The present article examines whether such bidirectional cross-linguistic influences (CLIs) exist at the phrase level.

Multiword expressions are heterogeneous, consisting of a large set of expression types, such as idioms (*kick the bucket*), lexical bundles (*in the middle of*), binomials (*bride and groom*), collocations (*strong tea*), and other phrasal elements (Siyanova-Chanturia and van Lancker Sidtis, 2019). MWEs vary greatly in frequency of occurrence<sup>1</sup>. However, what they have in common is that they are highly familiar and predictable to a native speaker (Siyanova-Chanturia and Martinez, 2015). For example, on hearing or reading the beginning of *fish and . . .*, a proficient language user is likely to complete it with the most likely word(s) *chips*. Due to their frequency and predictability, MWEs are processed faster than matched novel strings of language by L1 speakers (Arnon and Snider, 2010; Tremblay et al., 2011; Vilkaite, 2016) and L2 speakers (Jiang and Nekrasova, 2007; Siyanova-Chanturia et al., 2011; Hernández et al., 2016). Specifically, using priming paradigms, studies have found that the beginning of a MWE can prime its terminal word (Durrant and Doherty, 2010; Wolter and Gyllstad, 2011; Carrol and Conklin, 2014). For instance, in a primed lexical decision task, Wolter and Gyllstad (2011) observed *collocational* priming among L1 and L2 speakers for prime-target item pairs consisting of verb-noun collocations (*find job*) when compared with unrelated item pairs (*hear part*). Similarly, in an eye-tracking study, Carrol et al. (2016) found *idiom* priming effects in L1 and L2 reading, such that the final words in idioms (*spill the beans*) were skipped more often than the last words in control phrases (*drop the beans*).

## Evidence for Congruency Effect in MWE Processing in an L2

There has been growing interest in how congruency (i.e., similarity in form and meaning between the L1 and L2) may affect MWE processing in *bilinguals*. An L2 expression is congruent if it has a word-for-word translation equivalent form in the L1 (Conklin and Carrol, 2018). Cross-language overlap plays an important role in the L2 MWE processing; L2 speakers can process congruent MWEs more rapidly than incongruent L2-only MWEs (Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013; Carrol et al., 2016). Using a phrase-acceptability judgment task, Yamashita and Jiang (2010) found that lower proficiency Japanese English as a foreign language (EFL) learners made more errors with and responded more slowly to incongruent English-only (verb-noun and adjective-noun) collocations than to congruent collocations. Higher proficiency Japanese English as a second language (ESL) users also made more errors on incongruent collocations than on congruent ones, but they responded equally fast to the two types of collocations, indicating that proficiency may partially offset the effect of congruency. However, Wolter and Gyllstad (2011, 2013)

and Wolter and Yamashita (2018) observed that even high proficiency L2 speakers showed a robust processing advantage in response times for congruent vs. incongruent collocations. They observed that congruent collocations (verb-noun and adjective-noun) were processed significantly faster and more accurately than incongruent (English-only) collocations by advanced Swedish learners of English, in lexical decision (Wolter and Gyllstad, 2011) and acceptability judgment experiments (Wolter and Gyllstad, 2013). Comparable results were reported with Japanese-English bilinguals (Wolter and Yamashita, 2018). Similarly, L2 idiom processing studies also found a facilitative effect of congruency (Titone et al., 2015; Carrol et al., 2016). For example, in an eye-tracking study, Carrol et al. (2016) found that advanced Swedish learners of English processed congruent idioms faster than literal controls, whereas they processed incongruent (English-only) idioms and literal controls in a similar way (as indexed by the likelihood of skipping of the final word). Together, these studies suggest that congruency between languages can facilitate L2 MWE processing.

Although the above studies show a clear influence of L1 knowledge on the processing of L2 MWEs, evidence is mixed with regard to whether *unfamiliar* translated L1-only MWEs that do not exist in the L2 can also show a processing advantage over matched controls. L2 idiom studies have found that the L1 influence extends to the processing of translated L1-only items. Carrol and Conklin (2014), for example, found that high-proficiency Chinese-English bilinguals showed priming in a lexical decision task for translated Chinese-only idioms (e.g., *draw a snake and add . . . feet*) relative to matched controls (e.g., *draw a snake and add . . . hair*), whereas a control group of English monolinguals showed no priming. Similar findings were reported in a follow-up eye-tracking study with a similar population (Carrol and Conklin, 2017). Furthermore, in an eye-tracking study with highly proficient Swedish-English bilinguals, Carrol et al. (2016) directly compared facilitation for congruent idioms (e.g., *lose your head*) and Swedish-only idioms (e.g., *play monkey*) relative to literal controls (e.g., *hurt your head*, *taste monkey*, respectively). They found that translated Swedish-only idioms showed the same level of facilitation as did congruent idioms, and that there was no additional facilitatory effect for congruent idioms due to additional experience of the same combinations in the L2. These results suggest that the familiarity with L1 MWEs is a key driver of L2 idiom processing advantage, above and beyond L2 experience.

Conversely, studies on L2 *collocational* processing did not report a processing advantage for translated L1-only collocations compared to matched controls. In a study by Wolter and Yamashita (2015), two groups of Japanese-English bilinguals (intermediate and advanced) and English monolinguals completed a double lexical decision task, where they decided whether or not both words of a collocation, presented simultaneously, were real English words. Three types of items were used: translated Japanese-only collocations (*high effect*), English-only collocations (*busy road*), and non-collocations (*bad gift*). They found no processing advantage for translated Japanese-only collocations over non-collocations in either group of Japanese-English bilinguals, suggesting no activation

<sup>1</sup> Most MWEs are high frequency word combinations. Some, however, exhibit low frequencies (e.g., idioms and proverbs, see Moon, 1998).

of known L1 collocations in the processing of the translated L1-only items. In a follow-up study that encouraged focus on meaning rather than form, intermediate and advanced Japanese–English bilinguals and English monolinguals performed an acceptability judgment task, in which they decided as quickly as possible whether or not a two-word combination (*thick fog*) was commonly used in English (Wolter and Yamashita, 2018). Again, the results showed no significant processing advantage for translated Japanese-only collocations (*weak rain*) over non-collocations (*proud idea*).

## Mechanisms Behind Congruency Effect in MWE Processing in an L2

The issue of whether or not bilinguals show an advantage in the processing of L1-only MWEs translated into the L2 has important implications for understanding the mechanisms behind the congruency effect in the processing of L2 idioms and collocations and other MWEs (Wolter and Gyllstad, 2011, 2013; Carrol and Conklin, 2014, 2017; Yamashita, 2018). The L1 influence on L2 MWE processing may be explained in two ways. One explanation attributes the observed L1-on-L2 effect to the online activation of known L1 MWEs, i.e., the L1 MWE activation account (see also Yamashita, 2018; Zeng et al., 2020). The second explanation attributes the congruency advantage to the age of acquisition effect, assuming that congruent MWEs are acquired earlier and faster than L2-only MWEs. We will refer to this as the L2 MWE experience account.

In the L1 MWE activation account, known L1 MWEs are assumed to be automatically activated in L2 processing, leading to their faster processing (e.g., Wolter and Gyllstad, 2011; Carrol and Conklin, 2014; Carrol et al., 2016). For example, Carrol and Conklin (2014, concerning idioms) proposed that L2 words automatically activate L1 equivalents in bilinguals, which, in turn, trigger a known L1 sequence via direct retrieval of a unitary form. Likewise, concerning collocations, Wolter and Gyllstad (2011) proposed that an L2 word activates not only its L2 collocates (e.g., *strong* activates its collocate *tea*), but also its L1 translation equivalent (*strong* – *nong/浓*), which in turn activates its L1 collocates via collocational priming (*cha/茶* – *tea*). A number of studies have shown that when bilinguals process language in their L2, they obligatorily activate the L1 translation equivalents (i.e., cross-language translation priming; e.g., the L2 word *horse* primes its L1 translation equivalent *ma/马*) (Wu and Thierry, 2010; Zhang et al., 2011; Wu et al., 2013). For instance, in a relatedness judgment task with Chinese–English bilinguals, Wu and Thierry (2010) found the N400 effect for English word pairs whose Chinese translations had a repeated phonological component, e.g., *experience* [*Jing Yan* 经验]–*surprise* [*Jing Ya* 惊讶]. They concluded that L1 translations are automatically activated in L2 processing. Thus, it is plausible that due to cross-language activation in bilinguals, cross-language priming may extend to the phrase level. Under the L1 MWE activation account, when translation equivalents of L1 MWEs are first encountered in an L2, some facilitatory L1 influence in their processing should be observed (Carrol et al., 2016). This is supported by empirical studies that have reported on idiom priming effects for L1-only idioms over literal controls

in bilinguals when encountered in the L2 for the first time (Carrol and Conklin, 2014, 2017; Carrol et al., 2016). However, no support is found in the processing of translated L1-only collocations (Wolter and Yamashita, 2015, 2018).

In the L2 MWE experience account, no assumption is made about automatic activation of L1 MWE translation equivalents in L2 processing and, therefore, no priming for translated L1-only MWEs is predicted in L2 processing tasks. According to Wolter and colleagues (e.g., Wolter and Gyllstad, 2013; Wolter and Yamashita, 2015, 2018), congruent MWEs are acquired before incongruent MWEs due to positive L1 transfer, and, thus, congruent MWEs should be processed faster than incongruent MWEs. This would be analogous to the age-of-acquisition (AoA) effect, i.e., words that are acquired earlier are processed faster than words that are acquired later (Morrison and Ellis, 1995; Ellis and Morrison, 1998; Juhasz and Rayner, 2006). Multiple studies have shown that L1 plays an important role in the *acquisition* of L2 MWEs (Nesselhauf, 2005; Römer et al., 2014; Sonbul et al., 2020). For instance, Yamashita and Jiang (2010) found that both lower- and higher-proficiency Japanese–English bilinguals, but not monolingual English controls, made fewer errors on congruent collocations than incongruent L2-only collocations in a phrase-acceptability judgment task. They concluded that acquiring congruent L2 collocations takes less time and requires less exposure to the L2 than incongruent L2-only collocations. As Yamashita and Jiang (2010) posited, a congruent L2 MWE and its L1 counterpart share the identical or very similar concept, and thus bilinguals can easily accept and store congruent MWEs in memory by simply resorting to L1 expressions (Yamashita and Jiang, 2010, p. 662). Thus, it is plausible that congruent MWEs are acquired earlier than incongruent, L2-only MWEs.

With respect to the second claim by Wolter and colleagues (Wolter and Gyllstad, 2013; Wolter and Yamashita, 2015, 2018) that earlier acquired MWEs are processed faster than later acquired MWEs, there is empirical evidence showing that AoA affects the processing of units longer than a word. Using a phrasal decision task, Arnon et al. (2017) found that adults responded faster to early acquired phrases (*for the baby*) compared to late-acquired phrases (*for the teacher*), suggesting the AoA effect for units beyond single word level. Under this L2 MWE experience account, translated L1-only items which are encountered in the L2 for the first time should not show a processing advantage over matched controls in bilinguals. This view has found support in the results from Wolter and Yamashita (2015, 2018). Taken together, the L1 MWE activation account and L2 MWE experience account make differential predictions about the processing of translated L1-only MWEs when encountered in the L2 for the first time, although evidence is still mixed. Further research is needed to explore the processing of translated L1-only MWEs in bilinguals.

## L2 Influence on Lexical Processing in an L1

Additionally, although the reviewed studies have established that the L1 knowledge influences the processing of L2 MWEs, whether the processing of L1 MWEs is affected by the knowledge of L2

has not been sufficiently addressed in the literature. This issue, however, has been investigated in lexical, single word, processing research. The literature on the topic suggests that even weak, non-dominant L2 may affect the processing of words in the dominant L1 (van Hell and Dijkstra, 2002; Schoonbaert et al., 2009; Degani et al., 2011). For instance, van Hell and Dijkstra (2002) found that L1 words that are cognates with their L2 translations (e.g., Dutch–English: *bakker-baker*) lead to faster lexical decision responses than L1 non-cognate controls. In a study employing *non-cognate* translation pairs and a masked priming paradigm, Schoonbaert et al. (2009) demonstrated translation priming effects not only from L1 to L2 (*meisje-GIRL*), but also from L2 to L1 (*girl-MEISJE*). These studies suggest that lexical activation in bilingual memory operates in a parallel, language non-selective way, and that L1 processing can be influenced by the weaker L2, even when the task is completed exclusively in the L1 (Kroll and De Groot, 2009). Although cross-language influences have been reported in both directions, L1 typically has a higher impact on L2 processing than vice versa (Keatley et al., 1994; Jiang, 1999; Schoonbaert et al., 2009). Several cross-language priming studies have found strong priming from L1 to L2 and weaker or no priming from L2 to L1 (Gollan et al., 1997; Jiang, 1999; Finkbeiner et al., 2004; Schoonbaert et al., 2009).

The present study tests cross-linguistic influences in the processing of binomials (*knife and fork*) – a type of MWEs for which this issue has not yet been examined. We investigate whether L1 influences the processing of congruent L2 binomials and whether this influence extends to the processing of translated L1-only binomials (i.e., binomials which have not been previously seen in L2). Secondly, we test whether a bilingual's L2 influences the processing of binomials in the L1 and, if so, whether or not this influence is equally strong. We thus explore cross-language influences in both directions in the processing of binomials, addressing an important gap in MWE processing literature.

## THE PRESENT STUDY

To investigate cross-language influences in the processing of binomials, three groups of participants, Chinese–English and English–Chinese bilinguals and English monolinguals, completed the same English lexical decision experiment with a binomial priming manipulation. Both groups of bilinguals completed the experiment in the L2 immersion context. This design allowed us to investigate how the three groups of participants processed three types of MWEs: congruent, English-only, and translated Chinese-only binomials. In the case of Chinese–English bilinguals, we tested the involvement of L1 in L2 MWE processing, while with English–Chinese bilinguals, we tested the involvement of L2 in L1 MWE processing. The monolingual group of participants served as a baseline group.

Binomials are three-word phrases that are realized in English as an *A and B* form, where a specific word order is preferred (*knife and fork* vs. *fork and knife*) (Benor and Levy, 2006; Carrol and Conklin, 2020). They are highly fixed, that is, the reversed form is rarely used (Carrol and Conklin, 2020). The relative

frequency of “*A and B*” vis-à-vis the reversed form “*B and A*” is quite central to binomials, in that “*A and B*” is always more frequent than “*B and A*.” The experiment investigated whether the first two words of a binomial phrase facilitate lexical access to the final word of the phrase. The participants were briefly shown the first two words of a binomial and a control phrase (*knife + and* OR *spoon + and*) and then made lexical decisions on the final word (*fork*). We compared response times on the final words of binomials (*knife and fork*) and control items (*spoon and fork*). Shorter response times on the final word of the binomials compared to the controls (i.e., MWE priming) was taken as evidence that the binomial expressions were processed as highly familiar, conventional phrases. To test cross-language influences in the processing of binomials, we wanted to determine whether congruent binomials (i.e., English binomials whose Chinese translation equivalents are also binomials in Chinese) would be processed faster than English-only binomials (i.e., English binomials whose Chinese translation equivalents are not binomials in Chinese) by the bilinguals (but not by the monolinguals). This congruency effect is a prominent marker of cross-linguistic influences. We further sought to determine whether translated Chinese-only binomials (i.e., Chinese binomials whose English translation equivalents are not binomials) were processed faster than their control phrases, in order to better understand and interpret the mechanisms underpinning congruency effect in MWE processing.

The research questions we sought to answer are:

- (1) Is CLI observed in the processing of congruent L2 binomials by Chinese–English bilinguals?
- (2) Is CLI observed in the processing of congruent L1 binomials by English–Chinese bilinguals?
- (3) Is CLI observed in the processing of translated Chinese-only binomials with Chinese–English or English–Chinese bilinguals?

We predict that congruent binomials should be processed faster than English-only ones for Chinese–English and English–Chinese bilinguals. This is predicted by both the L1 MWE activation and L2 MWE experience accounts, although their proposed mechanisms responsible for the congruency advantage are different. We also predict that cross-linguistic influences should be greater in the L1-L2 direction than those in the L2-L1 direction, based on the findings reported in bilingual studies (Keatley et al., 1994; Jiang, 1999; Schoonbaert et al., 2009). In other words, we hypothesize that cross-language influences in the L2-L1 direction may occur but are likely to be weaker than those in the L1-L2 direction. In addition, we predict faster processing of translated Chinese-only binomials vs. controls in Chinese-English bilinguals, if the L1 MWE activation account is supported, or no MWE priming effect if the L2 MWE experience account is supported. English–Chinese bilinguals should show the same pattern in the processing of translated Chinese-only binomials as Chinese–English bilinguals, if their Chinese language proficiency is sufficiently high. Finally, we predict that English–Chinese bilinguals may process L1 MWEs in a different way from English monolinguals when they are in

an L2 immersion context, due to the need to inhibit interference from their L1, especially if their knowledge of L2 is comparatively weak. This prediction is based on studies on the influence of L2 immersion on L1 processing which have found that bilinguals immersed in an L2 environment show slower processing speed in L1 compared to those who have not experienced immersion (Linck et al., 2009; Baus et al., 2013; Morales et al., 2014).

## MATERIALS AND METHODS

### Participants

Three groups of participants were recruited for the study: Chinese–English bilinguals ( $n = 52$ ), English–Chinese bilinguals ( $n = 51$ ), and English monolingual controls ( $n = 52$ ). The number of participants was estimated based on a repeated measures design with the expected effect size being around  $d = 0.3$  for the power of 80% (Brysbaert and Stevens, 2018). Each participant received \$10 for their participation in the experiment. The study was conducted with the ethics approval from Victoria University of Wellington (VUW).

Chinese–English bilingual participants were undergraduate and postgraduate international students and young professionals studying or working at VUW. They completed a language background questionnaire before the experiment, in which they reported their English proficiency test score (International English Language Testing System [IELTS] or Test of English as a Foreign Language [TOEFL]), the number of years in an English-speaking country (average = 3.86 years, range: 0.5 – 17 years), and an estimate of their daily usage of English (average = 48%, range: 10% – 90%). Their mean IELTS score<sup>2</sup> was 6.67 (range: 6 – 8; roughly equivalent to the levels B2–C1 of the Common European Framework of Reference for Larson et al., 2014). They were thus regarded as advanced speakers of English as a second language.

English–Chinese bilingual participants were undergraduate and postgraduate students from Peking University and Tsinghua University, China. They were L1 English speakers who came to study Chinese or other subjects in Beijing as international students. They completed a language background questionnaire before the experiment, in which they reported their Chinese

proficiency<sup>3</sup> (self-reported), the number of years of exposure in China (average = 1.92 years, range: 0.6 – 8 years), and the estimation of their daily usage of Chinese (average = 37%, range: 5% – 90%). Twenty-two participants reported themselves as intermediate speakers of Chinese as a L2, and 29 participants as advanced speakers of Chinese as a L2.

English monolingual speakers were also undergraduate and postgraduate university students and young professionals, from VUW. They completed a language background questionnaire before the experiment to make sure they had no knowledge of Chinese. **Table 1** summarizes all participants' language proficiency characteristics.

### Materials

The critical materials consisted of 60 binomials and 60 control phrases. The binomials were of three types: (1) congruent binomials (e.g., *sun and moon*), (2) incongruent English-only binomials (e.g., *bread and butter*), and (3) translated Chinese-only binomials (e.g., *wisdom and strength*). Each binomial was paired with a control phrase. Control items were created by replacing the first word of the corresponding binomial with an alternative word that was semantically related to the final word of the binomial condition (e.g., *knife and fork* vs. *spoon and fork*). Binomials and their corresponding controls thus differed only in the first word. Control items formed semantically plausible low frequency phrases. This resulted in 120 experimental stimuli (60 binomials and 60 controls), see **Supplementary Appendix 1**. Examples of the materials for each condition are presented in **Table 2**.

### The Binomials and Their Phrase Frequency

The three types of binomials were chosen using the following criteria. First, for congruent binomials, the frequency of the binomial was much higher than the frequency of the reversed form in English and Chinese. For example, the binomial, *sun and moon* (太阳和月亮, taiyang he yueliang), is much more frequent than the reversed form, *moon and sun* (月亮和太阳, yueliang he taiyang), in English and Chinese: 30.54 vs. 6.25 occurrences (per 100 million words) in the Corpus of Contemporary American English (COCA: 560 million words) (Davies, 2008), and 38.63 vs. 7.77 occurrences (per 100 million words) according to the corpus of Center for Chinese Linguistics Peking University<sup>4</sup>

<sup>2</sup> Among the 52 Chinese–English bilingual participants, there were 39 participants with IELTS scores, 4 participants with TOEFL scores, and 9 participants who had no English proficiency test record but took the University's English Proficiency Program (EPP) and met their English proficiency requirement. TOEFL scores were converted to IELTS band scores for the ease of comparison based on the Comparison Table provided on the website: <https://www.ets.org/toefl/institutions/scores/compare/>.

<sup>3</sup> Only nine participants had taken the standardized Chinese proficiency test called Hanyun Shuiping Kaoshi (HSK), which was not compulsory for their programs. Therefore, we used self-reported measures to assess their Chinese proficiency.

<sup>4</sup> The CCL corpus contains 700 million Chinese characters. However, the character number is not equivalent to word number (e.g., 东西 dong xi, meaning 'thing(s)' in

**TABLE 1 |** Means (standard deviations) of self-reported age, L2 proficiency levels, daily usage of L2, years of exposure to L2 in L2-speaking countries.

|                         | Chinese–English ( $N = 52$ ) | English–Chinese ( $N = 51$ ) | English monolinguals ( $N = 52$ ) |
|-------------------------|------------------------------|------------------------------|-----------------------------------|
| Age                     | 28.46 (6.16)                 | 22.88 (2.85)                 | 23.85 (6.04)                      |
| English proficiency     | Advanced                     | Native                       | Native                            |
| Chinese proficiency     | Native                       | Intermediate+                | 0.00 (0.00)                       |
| Daily usage of L2       | English: 48% (23%)           | Chinese: 37% (22%)           | N/A                               |
| Years of exposure to L2 | 3.86 (3.88)                  | 1.92 (1.84)                  | N/A                               |

**TABLE 2 |** Example of stimulus materials for each condition.

| Condition    | Binomial                   | Control                      |
|--------------|----------------------------|------------------------------|
| Congruent    | Sun and <u>moon</u>        | Star and <u>moon</u>         |
| English-only | Bread and <u>butter</u>    | Toast and <u>butter</u>      |
| Chinese-only | Wisdom and <u>strength</u> | Exercise and <u>strength</u> |

(CCL: 437.5 million words, Zhan et al., 2003). It can thus be classified as a true binomial both in English and Chinese. Congruent binomials and their reversed forms differed in phrase frequency in English (binomials: mean = 69.54,  $SD = 90.89$ ; reversed forms: mean = 8.63,  $SD = 9.46$ ;  $t = 6.95$ ,  $p < 0.0001$ ) and Chinese (binomials: mean = 66.21,  $SD = 75.41$ ; reversed forms: mean = 6.57,  $SD = 9.38$ ;  $t = 6.97$ ,  $p < 0.0001$ ).

Second, for English-only binomials, the frequency of the binomial was higher than the frequency of the reversed form in *English* but not in Chinese. The combination was legal in Chinese, but there was no word order preference in terms of frequency of occurrence. For example, the English binomial *bread and butter* is more frequent than the reversed form *butter and bread* (71.79 vs. 2.32 occurrences in COCA). However, the Chinese translation equivalent for the binomial *bread and butter*, 面包和黄油 (mianbao he huangyou), is almost as frequent as that of the reversed form *butter and bread*, 黄油和面包 (huangyou he mianbao): 3.89 vs. 1.83 occurrences in CCL. It is therefore classified as an English-only binomial. English-only binomials differed from their reversed forms significantly in phrase frequency (binomials: mean = 106.90,  $SD = 217.56$ ; reversed forms: mean = 6.28,  $SD = 10.10$ ;  $t = 7.20$ ,  $p < 0.0001$ ), whereas their Chinese translation equivalents were as frequent as their reversed forms (binomials: mean = 2.07,  $SD = 2.11$ ; reversed forms: mean = 1.12,  $SD = 1.10$ ;  $t = 1.66$ ,  $p = 0.11$ ). Additionally, to ensure the difference in the processing of congruent and English-only binomials could be attributed to the difference in congruency rather than phrase frequency, we also matched congruent and English-only binomials for phrase frequency in English (congruent binomials: mean = 69.54,  $SD = 90.89$ ; English-only binomials: mean = 106.90,  $SD = 217.56$ ;  $t = -0.08$ ,  $p = 0.94$ ).

Third, for Chinese-only binomials, the frequency of the binomial had to be higher than the phrase frequency of the reversed form in *Chinese* but not English. That is, for Chinese-only binomials there was no word-order preference in English. For example, the Chinese binomial 智慧和力量 (zhihui he liliang, wisdom and strength) was much more frequent than the reversed form 力量和智慧 (liliang he zhihui, strength and wisdom): 132.34 vs. 22.4 occurrences in CCL. By contrast, the English translation equivalent for the binomial 智慧和力量, *wisdom and strength*, was almost as frequent as that of the reversed form 力量和智慧, *strength and wisdom*: 3.04 vs. 4.64 occurrences in COCA. It was thus regarded as a Chinese-only binomial. Chinese-only binomials and their reversed forms differed in phrase frequency

English, is considered as a two-character word). Following the Lancaster Corpus of Mandarin Chinese (McEnery and Xiao, 2004), we adopted a ratio of 1:1.6 between words and characters to calculate the number of words in CCL, which amounts to 437.5 million.

in Chinese (binomials: mean = 213.17,  $SD = 353.71$ ; reversed forms: mean = 7.09,  $SD = 11.72$ ;  $t = 8.98$ ,  $p < 0.0001$ , but not in English (binomials: mean = 1.97,  $SD = 2.12$ ; reversed forms: mean = 1.53,  $SD = 1.60$ ;  $t = 0.50$ ,  $p = 0.62$ ).

Unlike English binomials which have a fixed structure of *A and B*, Chinese binomials are more flexible in form, in that they can take the following three forms: *A and B*, *AB*, and *A, B* (e.g., *knife and fork*: 刀和叉, 刀叉, 刀、叉). This reflects the characteristics of Chinese language, which is a paratactic language, whereby connective elements are often optional or unnecessary (Li and Ho, 2016). For binomials in Chinese, the word order is the most important attribute (i.e., *A* precedes *B*, rather than *B* precedes *A*), while the coordinator is not necessary. Thus, when we identified the frequency of occurrence of a Chinese phrase in CCL, we extracted its frequency in the forms of *A and B*, *AB*, and *A, B*, and used the sum of their frequency as the frequency of occurrence of this phrase. The controls, however, always had a conjunction (e.g., 和 he: and) in the Chinese version. In addition, when we translated Chinese-only binomials to English, the addition of the conjunction 'and' was necessary to conform to the *A and B* structure of English binomials. This kind of variation in form due to language differences is often inevitable in cross-language studies (e.g., Carrol and Conklin, 2014; Carrol et al., 2016).

Most of the binomials used in our study are literal phrases. However, in the congruent category, two items have a figurative and a literal meaning ('song and dance,' 'thick and thin'). In the English-only category, three items have both a figurative and a literal meaning ('bread and butter,' 'sticks and stones,' and 'bed and breakfast'). Therefore, literality was comparable across the different lists of binomials.

### Association Strength

Following Siyanova-Chanturia et al. (2011), the University of South Florida (USF) Free Association Norms database<sup>5</sup> was used to match the constituents (i.e., the first content word and the second content word) of the binomials (*sun and moon*) and the control items (*star and moon*) in forward association strength (*sun-moon* vs. *star-moon*: 0.15 vs. 0.115). This was needed to ensure that any processing advantage for binomials over their corresponding controls was not due to the first word in the binomials (*sun*) being a better prime than the first word in the control items (*star*) for the same target (*moon*) (e.g., Siyanova-Chanturia et al., 2011). There was no significant difference in forward association strength between the components of congruent and English-only binomials and their corresponding controls (congruent condition,  $t = 1.41$ ,  $p = 0.17$ ; English-only condition,  $t = 1.75$ ,  $p = 0.15$ ). However, for binomials and their controls in the Chinese-only category, the association strength between their constituents was not attested in the USF norm database. This was expected, since the USF is based on English, while no comparable Chinese database exists for the Chinese language. We only included the items which existed in the database.

<sup>5</sup><http://w3.usf.edu/FreeAssociation/>

## Word Length and Frequency of the First Content Word

The first words in the binomial and control conditions were matched for part of speech, word length, and frequency (where possible). There was no significant difference between the first words in the binomial and the control conditions for word length (congruent condition,  $t = -1.19$ ,  $p = 0.24$ ; English-only condition,  $t = -1.04$ ,  $p = 0.28$ ; Chinese-only condition,  $t = -0.07$ ,  $p = 0.94$ ). However, while the first words in congruent and Chinese-only binomials and their corresponding control phrases were matched in terms of lexical frequency (congruent condition,  $t = 1.24$ ,  $p = 0.23$ ; Chinese-only condition,  $t = 0.52$ ,  $p = 0.61$ ), the first words in English-only binomials and their matched controls could not be matched ( $t = 3.01$ ,  $p = 0.005$ ). It was impossible to create plausible control items matched in frequency as well as forward association strength. To partial out any possible effect of the first word's lexical frequency, we added the frequency of the first word as a covariate in our initial statistical model. The properties of the experimental items are presented in **Table 3**.

## Fillers and Non-word Items

A set of fillers with the same syntactic structure as binomials was constructed to reduce the proportion of related prime-target pairs, following 1/5 ratio proposed by McNamara (2005). The fillers were grammatical but implausible (*business and soul*). Non-word items were created to make an equal number of word/non-word responses, with the syntactic structure of *word + and + non-word*. All non-words came from the ARC non-word database (Rastle et al., 2002). They conformed to the phonotactic rules of English and were matched with the other items for length (mean = 5.88 letters). Primes for the non-word targets were words that were not used in other conditions. See **Supplementary Appendix 2** for fillers and non-word items used in the experiment.

## Design

A repeated-measures design was used, with each participant exposed to the critical items in both conditions; this allowed for a within-participant comparison of response times in the two experimental conditions, providing better control for individual differences (Millar, 2011). To control for the repetition effect, two counterbalanced presentation lists were constructed. Half of

the critical targets per list were presented as binomials and half as control phrases. In addition, the numbers of stimuli of each congruency type in each list were also balanced, such that each list contained an equal number of congruent, English-only, and Chinese-only binomials. Participants were randomly assigned to one of the two groups in the order of their participation. Group 1 saw List A first and then List B, and for Group 2 the order was reversed. The same number of participants was assigned to each group.

## Procedure

The experiment was conducted in a laboratory using DMDX software (Forster and Forster, 2003). Participants first read instructions on the computer screen and then completed 20 practice trials. All items were presented in the middle of the screen in white lowercase letters in Courier New font, size 24 pt, over a black background. At the start of each trial, a fixation point (“++++”) was presented in the middle of the screen for 500 ms. It was replaced with the first word prime (*knife* in “*knife and fork*”), which was displayed for 250 ms. After that, a blank screen was presented for 150 ms (inter-stimulus interval [ISI] = 150 ms). Then the second word prime “*and*” was displayed for 250 ms, followed by the same ISI (150 ms). Finally, the target appeared and remained on the screen until a response was made, or the item timed out at 3,000 ms. The procedure is summarized in the following diagram:



The items were presented in two counterbalanced blocks of 154 trials, with a self-paced break after Block 1. Within each block, the trial order was randomized for each participant. The whole experiment took approximately 20 minutes to complete.

## ANALYSIS AND RESULTS

We analyzed accuracy and response latencies (RT). In the accuracy analysis, all responses were included. The mean response accuracy to non-word items was 95.94% for English monolinguals, 96.54% for English–Chinese bilinguals, and 82.18% for Chinese–English bilinguals. Accuracy was coded as a

**TABLE 3 |** Means (standard deviations) of phrase frequency, word length and frequency of first word, and semantic association strength for the binomial and control conditions (counts based on occurrences per 100 million words).

|                                   | Congruent            |                        | English-only         |                      | Chinese-only         |                         |
|-----------------------------------|----------------------|------------------------|----------------------|----------------------|----------------------|-------------------------|
|                                   | Binomial             | Control                | Binomial             | Control              | Binomial             | Control                 |
| Phrase frequency (English corpus) | 69.54(90.89)         | 0.95 (1.03)            | 106.90(217.56)       | 0.76(0.94)           | 1.97(2.12)           | 0.52(0.57)              |
| (Chinese corpus)                  | 66.21(75.41)         | 0.55(1.21)             | 2.07(2.11)           | 0.13(0.23)           | 213.17(353.71)       | 0.56(1.28)              |
| First word length                 | 5.45 (1.70)          | 6.05 (1.70)            | 4.8 (1.28)           | 5.3 (1.56)           | 6.4 (2.28)           | 6.55(2.48)              |
| First word frequency              | 7636.60<br>(7400.19) | 15633.22<br>(46006.48) | 8559.96<br>(8485.87) | 4544.17<br>(7898.84) | 5700.86<br>(5918.46) | 5698.47<br>(9866.77)    |
| Association strength              | 0.24 (0.22)          | 0.17 (0.15)            | 0.14 (0.17)          | 0.07 (0.09)          | 0.01(0.01)           | 0.01(0.03) <sup>a</sup> |

<sup>a</sup>For items in the Chinese-only category, we could only obtain the association strength between the constituents of six binomials and four control phrases. The values reported were based on these 10 items which existed in USF norm database.

binary variable (1 – correct; 0 – incorrect). A generalized linear mixed-effects regression analysis was conducted to compare the accuracy between language groups. The likelihood ratio test indicated that there was a significant difference in response accuracy between the language groups ( $\chi^2 = 76.33$ ,  $p < 0.0001$ ). To further explore these differences, *post hoc* tests were run for the significant interactions, using `emmeans()` function in the R package `emmeans` (Lenth, 2019), with Bonferroni adjustments. The results showed that there was no significant difference in terms of accuracy for non-words between English monolinguals and English–Chinese bilinguals ( $z = -0.86$ ,  $p = 0.39$ ). However, Chinese–English bilinguals had lower accuracy for non-word trials than English monolinguals ( $z = 8.18$ ,  $p < 0.0001$ ) and English–Chinese bilinguals ( $z = 8.95$ ,  $p < 0.0001$ ). On word trials, the mean accuracy was 98.39% for English monolinguals, 97.68% for English–Chinese bilinguals, and 97.27% for Chinese–English bilinguals. Pairwise comparisons showed a significant difference in response accuracy between English monolinguals and Chinese–English bilinguals ( $z = 3.21$ ,  $p = 0.004$ ), but not between English monolinguals and English–Chinese bilinguals ( $z = 2.04$ ,  $p = 0.12$ ), or between Chinese–English and English–Chinese bilinguals ( $z = -1.17$ ,  $p = 0.73$ ). Importantly, within each language group, there was no significant difference in response accuracy between the binomial and control conditions for the three congruency types (i.e., congruent, English-only and Chinese-only). That is, there was no response accuracy priming for any of the three types of binomials in any language group.

For the RT analyses, the data for non-word and filler items were excluded from the analysis. We performed the analyses on RTs to 60 binomials (20 items for each congruency type: congruent, English-only, Chinese-only) and their corresponding controls (120 items in total). Incorrect responses were removed from the RT analysis, resulting in the loss of 1.12% data for English monolinguals, 2.31% data for Chinese–English bilinguals, and 1.98% data for English–Chinese bilinguals. Extreme values (RTs longer than 2000 ms or shorter than 250 ms) were also excluded (e.g., Sprenger et al., 2006; Matsuno, 2017), which resulted in the loss of 0.14% data for English monolinguals, 0.56% data for Chinese–English bilinguals, and 0.17% data for English–Chinese bilinguals.

English monolinguals overall responded faster than Chinese–English bilinguals (monolinguals: mean = 497 ms,  $SD = 133$ ; bilinguals: mean = 655 ms;  $SD = 219$ ). There was a difference of about 150 ms in RTs on the targets between English monolinguals and Chinese–English bilinguals, which was consistent with previous studies (e.g., Jiang, 1999). The mean RT to the target words for English–Chinese bilinguals was 512 ms ( $SD = 147$ ), which was 15 ms slower than that for English monolinguals. Means of RTs by condition for three groups of participants is shown in **Table 4**.

Following Carrol and Conklin (2014) and Wolter and Yamashita (2015), reaction time data for each group were analyzed separately with linear mixed effects model using R (R Core Team, 2016), using `lme4` package (Bates et al., 2015), and `lmerTest` package (Kuznetsova et al., 2015). Following Brysbaert and Stevens (2018), RTs were inverse transformed (i.e.,  $-1000/RT$ ) to bring the data closer to normal distribution.

**TABLE 4 |** Descriptive statistics: mean response times in ms (standard deviations) and difference between mean response times to the binomial and control phrases for English monolinguals, Chinese–English, and English–Chinese bilinguals in each of the six experimental conditions.

|                            |              | Binomial        | Control         | Difference |
|----------------------------|--------------|-----------------|-----------------|------------|
| English monolinguals       | Congruent    | 479.42 (125.88) | 505.11 (134.61) | 25.69      |
|                            | English-only | 471.07 (132.57) | 492.68 (129.82) | 21.63      |
|                            | Chinese-only | 519.12 (138.14) | 513.56 (128.80) | –5.56      |
| Chinese–English bilinguals | Congruent    | 637.84 (221.73) | 657.86 (213.51) | 20.02      |
|                            | English-only | 641.41 (204.58) | 645.09 (205.91) | 3.68       |
|                            | Chinese-only | 676.97 (241.64) | 669.25 (222.02) | –7.72      |
| English–Chinese bilinguals | Congruent    | 495.43(128.65)  | 516.26(14.013)  | 20.83      |
|                            | English-only | 501.56(152.66)  | 504.96(141.42)  | 3.4        |
|                            | Chinese-only | 528.79(158.83)  | 523.23(152.79)  | –5.56      |

Difference is calculated with mean RT to controls minus mean RT to binomials.

Inverse-transformed RTs were normally distributed, with skewness of 0.07 and kurtosis of 2.88. Inverse-transformed RT to the final word of each phrase was used as the response variable.

For each group, the model fitting procedure started with the same maximal model with participants and items treated as random-effect factors. The following predictors were included: (1) item type (binomial vs. control), (2) congruency (congruent vs. English-only vs. Chinese-only), (3) English phrase frequency (counts based on occurrences per 100 million words, log transformed), (4) the frequency of the first content word of a phrase (counts based on occurrences per 100 million words, log transformed), and (5) forward association strength between the first word and the last word of a phrase (based on USF database, log transformed). Block order (order in which participants saw the two presentation lists: Order 1 vs. Order 2) and the trial number of the presentation of the phrase in the experiment (scaled) were considered as fixed effects to account for repetition priming and the longitudinal effect of the experimental task on the behavior of the participants. The model included the following interactions: (1) item type and congruency, (2) item type and phrase frequency, (3) item type and the frequency of the first content word, (4) congruency and phrase frequency, and (5) association strength and congruency. Starting with the maximal model, we used `step()` function in `lmerTest` to arrive at the best model fit. The initial model with random slopes failed to converge, so we did not include random slopes at this stage. After fitting the best model, we conducted a forward stepwise model selection to identify the appropriate random effects structure with random slopes, using Akaike information criterion (AIC) values.

In order to address the issue of the collinearity between phrase frequency and item type, we orthogonalized phrase frequency by fitting a linear model in which phrase frequency was predicted by item type, following Siyanova-Chanturia et al. (2011). The residuals of this model (`EngPhrFreq.Residual`) were then used as our predictor of phrase frequency, such that effects of item type were partialled out.

After identifying the best model with random slopes, we visually inspected a quantile–quantile plot of the model's

**TABLE 5 |** Results of mixed model for English monolinguals.

| Fixed effect                             | Estimate        | Standard error | df    | t      | p         |
|------------------------------------------|-----------------|----------------|-------|--------|-----------|
| Intercept                                | -1.98           | 0.04           | 97.86 | -45.14 | <2.00e-16 |
| ItemTypebinomial                         | -0.10           | 0.02           | 5563  | -5.58  | 2.59e-08  |
| Congruency (C-only)                      | 0.03            | 0.03           | 82.24 | 0.11   | 0.92      |
| Congruency (E-only)                      | -0.06           | 0.03           | 76.85 | -1.93  | 0.05      |
| EngPhrFreq.Resid                         | -0.01           | 0.01           | 1308  | -1.12  | 0.26      |
| AssoStrength.log                         | -0.37           | 0.07           | 628.9 | -5.01  | 6.96e-07  |
| TrialNum.sc                              | -0.05           | 0.01           | 50.84 | -4.41  | 5.43e-05  |
| BlockOrder2                              | -0.13           | 0.05           | 50.08 | -2.66  | 0.01      |
| ItemType (binomial) *Congruency (C-only) | 0.08            | 0.03           | 2963  | 2.45   | 0.01      |
| ItemType (binomial) *Congruency (E-only) | -0.01           | 0.02           | 5921  | -0.46  | 0.65      |
| ItemType (control) * Wrd1Freq.log.c      | 0.01            | 0.01           | 1386  | 2.25   | 0.02      |
| ItemType (binomial) * Wrd1Freq.log.c     | 0.04            | 0.01           | 1546  | 4.87   | 1.26e-06  |
| <b>Random effects</b>                    | <b>Variance</b> | <b>SD</b>      |       |        |           |
| Target                                   | 0.008           | 0.09           |       |        |           |
| Participant                              | 0.03            | 0.18           |       |        |           |
| TrialNum.sc   Participant                | 0.005           | 0.07           |       |        |           |
| Residual                                 | 0.12            | 0.35           |       |        |           |

df, degrees of freedom; Intercept levels: ItemType, control; Congruency, Congruent.  
Marginal  $R^2 = 0.08$ , Conditional  $R^2 = 0.33$ .

**TABLE 6 |** Results of mixed model for Chinese–English bilinguals.

| Fixed effect                             | Estimate        | Standard error | df     | t      | p         |
|------------------------------------------|-----------------|----------------|--------|--------|-----------|
| Intercept                                | -1.61           | 0.05           | 118.70 | -29.81 | <2.00e-16 |
| ItemTypebinomial                         | -0.05           | 0.02           | 159.50 | -3.39  | 0.0009    |
| Congruency (C-only)                      | 0.01            | 0.06           | 61.70  | 0.09   | 0.93      |
| Congruency (E-only)                      | -0.06           | 0.06           | 58.76  | -1.01  | 0.32      |
| AssoStrength.log                         | -0.18           | 0.08           | 139.03 | -2.27  | 0.02      |
| TrialNum.sc                              | -0.08           | 0.01           | 52.99  | -7.53  | 6.45e-10  |
| ItemType (binomial) *Congruency (C-only) | 0.06            | 0.02           | 58.92  | 2.38   | 0.02      |
| ItemType (binomial) *Congruency (E-only) | 0.06            | 0.02           | 186.20 | 2.63   | 0.009     |
| <b>Random effects</b>                    | <b>Variance</b> | <b>SD</b>      |        |        |           |
| Target                                   | 0.03            | 0.18           |        |        |           |
| EngPhrFreq.Resid   Target                | 0.001           | 0.03           |        |        |           |
| TrialNum.sc   Target                     | 0.0001          | 0.01           |        |        |           |
| Participant                              | 0.06            | 0.24           |        |        |           |
| TrialNum.sc   Participant                | 0.004           | 0.07           |        |        |           |
| Residual                                 | 0.08            | 0.28           |        |        |           |

df, degrees of freedom; Intercept levels: ItemType, control; Congruency, Congruent.  
Marginal  $R^2 = 0.04$ , Conditional  $R^2 = 0.57$ .

residuals and removed 2.5 SD from the residuals to satisfy the assumption of homoscedasticity and normal distribution (data loss: 1.83% for English monolinguals; 1.85% for Chinese–English bilinguals, 1.87% data for English–Chinese bilinguals). We refit the model with the new data. The results for the identified model are shown in **Table 5** for English monolinguals, **Table 6** for Chinese–English bilinguals, and **Table 7** for English–Chinese bilinguals.

## Results for English Monolinguals

The final model for English monolinguals included two significant two-way interactions (item type  $\times$  congruency, item

type  $\times$  Word 1 frequency). There were also statistically significant main effects of association strength, block order and trial number. The results suggested that more strongly associated phrases had overall shorter response latencies. Also, participants responded faster in Block 2 than in Block 1. They also went faster as the number of trials increased. The two-way interaction between item type and congruency ( $F = 3.82$ ,  $p = 0.02$ ) (**Figure 1**) showed that the English monolingual speakers processed congruent and English-only binomials significantly faster than their corresponding controls (i.e., priming effects are observed for congruent and English-only binomials), but there was no difference between Chinese-only binomials and

**TABLE 7** | Results of mixed model for English–Chinese bilinguals.

| Fixed effect                             | Estimate        | Standard error | df     | t      | p         |
|------------------------------------------|-----------------|----------------|--------|--------|-----------|
| Intercept                                | −2.01           | 0.04           | 99.99  | −48.47 | <2.00e-16 |
| ItemTypebinomial                         | −0.05           | 0.02           | 130.49 | −2.72  | 0.007     |
| Congruency (C-only)                      | −0.04           | 0.03           | 84.37  | −1.36  | 0.18      |
| Congruency (E-only)                      | −0.07           | 0.03           | 77.60  | −2.59  | 0.01      |
| EngPhrFreq.Resid                         | −0.03           | 0.01           | 177.49 | −2.74  | 0.006     |
| Wrd1Freq.log.c                           | 0.01            | 0.01           | 40.61  | 0.77   | 0.44      |
| AssoStrength.log                         | −0.25           | 0.08           | 161.17 | −3.21  | 0.002     |
| TrialNum.sc                              | −0.04           | 0.01           | 51.70  | −3.51  | 0.0009    |
| ItemType (binomial) *Congruency (C-only) | 0.02            | 0.04           | 173.83 | 0.55   | 0.59      |
| ItemType (binomial) *Congruency (E-only) | 0.06            | 0.02           | 179.42 | 2.32   | 0.02      |
| ItemTypebinomial: wrd1Freq.log.c         | 0.02            | 0.01           | 176.52 | 2.08   | 0.04      |
| <b>Random effects</b>                    | <b>Variance</b> | <b>SD</b>      |        |        |           |
| Target                                   | 0.007           | 0.08           |        |        |           |
| TrialNum.sc   Target                     | 0.0003          | 0.02           |        |        |           |
| Wrd1Freq.log.c                           | 0.0009          | 0.03           |        |        |           |
| Participant                              | 0.06            | 0.24           |        |        |           |
| TrialNum.sc   Participant                | 0.004           | 0.07           |        |        |           |
| Residual                                 | 0.11            | 0.33           |        |        |           |

df, degrees of freedom; Intercept levels: ItemType, control; Congruency, congruent. Marginal  $R^2 = 0.02$ , Conditional  $R^2 = 0.41$ .

their controls (i.e., no priming for Chinese-only). It also showed that priming effects for congruent and English-only binomials were comparable. To further explore these differences, *post hoc* tests were run for the significant interactions using `emmeans()` function in the R package `emmeans` (Lenth, 2019), with Bonferroni adjustments. The result is shown in **Table 8**.

For the English monolinguals, priming effect was observed for the congruent ( $t = 5.73$ ,  $p < 0.0001$ ) and English-only conditions ( $t = 5.94$ ,  $p < 0.0001$ ). The mean RT<sup>6</sup> to the binomials was 22 ms faster than RT to the control items (459 vs. 481 ms) in the congruent condition and 23 ms faster in the English-only condition (444 vs. 467 ms) (**Table 8**). We did not find priming for the Chinese-only condition ( $t < 1$ ,  $p = 0.99$ ), which confirmed that Chinese-only items were not processed by the English monolinguals as binomials.

## Results for Chinese–English Bilinguals

The final model for Chinese–English bilinguals revealed a significant interaction between item type and congruency ( $F = 4.33$ ,  $p = 0.016$ ). There were also statistically significant main effects of association strength and trial number. Words within more strongly associated phrases had overall shorter response latencies. As the number of trials increased, the response time became faster. The two-way interaction (**Figure 2**) showed that the Chinese–English bilinguals processed congruent binomials significantly faster than the controls, but there was no difference between their processing of the English-only binomials vs. controls, nor any difference between Chinese-only binomials vs. controls. That is, only congruent binomials showed

a priming effect. To further explore these differences, *post hoc* tests were run for the significant interaction using `emmeans()` function, with Bonferroni adjustments. The result is shown in **Table 9**.

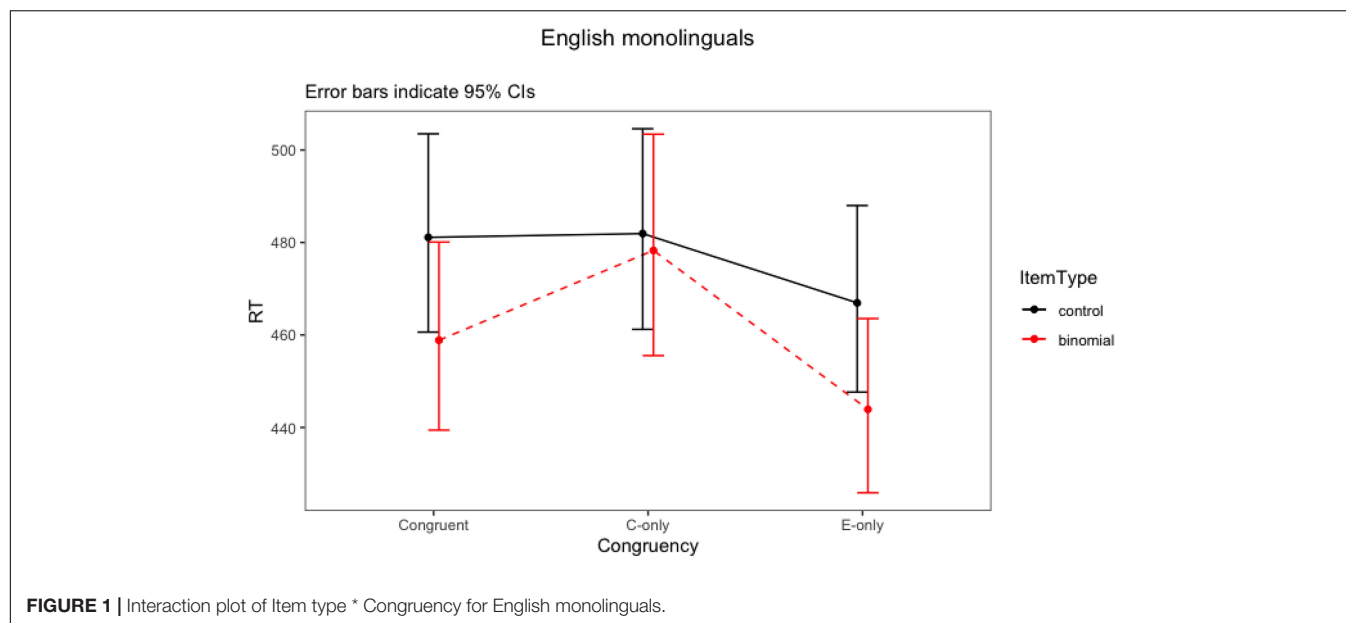
For Chinese–English bilinguals, the priming effect was only observed in the congruent condition ( $t = 3.39$ ,  $p = 0.01$ ), with RT to the terminal word in the binomials 19 ms faster than RT to the control items (595 vs. 614 ms, respectively). No priming effect was present for English-only binomials ( $t < 1$ ,  $p = 0.99$ ) nor Chinese-only ( $t < 1$ ,  $p = 0.99$ ) binomials: there was no difference between the binomials and the control items in English-only condition (594 vs. 593 ms) nor in the Chinese-only condition (618 vs. 616 ms). This suggests that only congruent expressions were processed as binomials, whereas English-only and Chinese-only items were not.

In sum, the relative processing advantage for congruent over English-only binomials compared to their corresponding controls was found for the Chinese–English participants, whereas for the monolingual participants no such difference was observed. In other words, the congruent binomials had a processing advantage over the English-only binomials for the Chinese–English bilinguals, even though the two types of binomials had been matched in phrase frequency.

## Results for English–Chinese Bilinguals

The final model for English–Chinese bilinguals revealed a trend toward an interaction between item type and congruency ( $F = 2.73$ ,  $p = 0.07$ ). There were also statistically significant main effects of English phrase frequency, association strength and trial number. The model suggested that phrase frequency was always facilitative (led to lower overall RTs). Association strength was also facilitative whereby more strongly associated phrases led to lower overall RTs. Participants responded faster as

<sup>6</sup>Here and the following reported in-text are model estimates. We report descriptive statistics calculated prior to the data analyses in **Table 4**.



the trial number increased. The two-way interaction (**Figure 3**) showed that the English-Chinese bilinguals processed congruent binomials somewhat faster than the controls, but there was no difference between their processing of the English-only binomials vs. controls, nor any difference between Chinese-only binomials vs. controls. To further explore these differences, *post hoc* tests were run for the significant interactions using *emmeans()* function, with Bonferroni adjustments. The result is shown in **Table 10**.

For English-Chinese bilinguals, there was a significant priming trend for congruent binomials. The magnitude of the priming effect was 12 ms (model estimate). However, it did not reach statistical significance, after a correction for multiple comparisons had been applied ( $t = 2.72$ ,  $p = 0.11$ ). There was therefore a *weak* priming effect for congruent binomials. In addition, no priming effects was observed in English-only ( $t < 1$ ,  $p = 0.99$ ) or Chinese-only ( $t = 1.19$ ,  $p = 0.99$ ) conditions.

## DISCUSSION

The aim of the present study was to examine whether there were CLIs in the processing of MWEs, in the direction of L1-L2 as well as in the reverse direction, L2-L1. We focused

on binomial expressions, that is, literal and compositional formulaic sequences, which have so far received little attention in cross-language processing research. To this end, we used a primed lexical decision task to examine the processing of three types of binomials vs. their corresponding matched infrequent controls: congruent English-Chinese (*sun and moon* vs. *star and moon*), English-only (*bread and butter* vs. *toast and butter*) and translated Chinese-only (*wisdom and strength* vs. *exercise and strength*). Three groups of participants, Chinese-English, English-Chinese bilinguals and English monolinguals, were tested.

## Cross-Language Influences From L1 to L2

English monolingual participants showed significant facilitation in the processing of the final word in English binomials compared to control phrases. The facilitation was irrespective of congruency; the magnitude of the priming effect was comparable in the congruent (22 ms) and English-only (23 ms) conditions. This offers further support to the tenet that binomials are processed differently from novel controls (Siyanova-Chanturia et al., 2011, 2017).

The Chinese-English bilinguals showed a priming effect (19 ms) in the processing of the terminal words of the congruent binomials compared to the control phrases, but no priming was observed for the English-only binomials (−1 ms). This indicates that there was a processing advantage for the congruent L2 binomials over the English-only binomials for the Chinese-English bilinguals (cf. RQ1). This result is in line with previous studies involving other types of MWEs. For example, in Wolter and Gyllstad (2011, 2013) and Carrol et al. (2016), bilinguals (but not monolinguals) showed a congruency advantage in the processing of congruent L2-L1 over L2-only idioms and collocations, respectively. Since congruent

**TABLE 8 |** Results of *post hoc*, within-group tests of RTs for congruent, English-only, and Chinese-only items relative to the control items for English monolinguals.

| Contrast    | Group | Estimate | Standard error | t     | p       | ED (msec) |
|-------------|-------|----------|----------------|-------|---------|-----------|
| Ctrl-Cngr   | ENS   | 0.101    | 0.018          | 5.573 | <0.0001 | 22        |
| Ctrl-E only | ENS   | 0.111    | 0.019          | 5.944 | <0.0001 | 23        |
| Ctrl-C only | ENS   | 0.016    | 0.025          | 0.645 | 0.99    | 4         |

ED (estimated difference) is calculated with RTs to controls minus RTs to binomials.

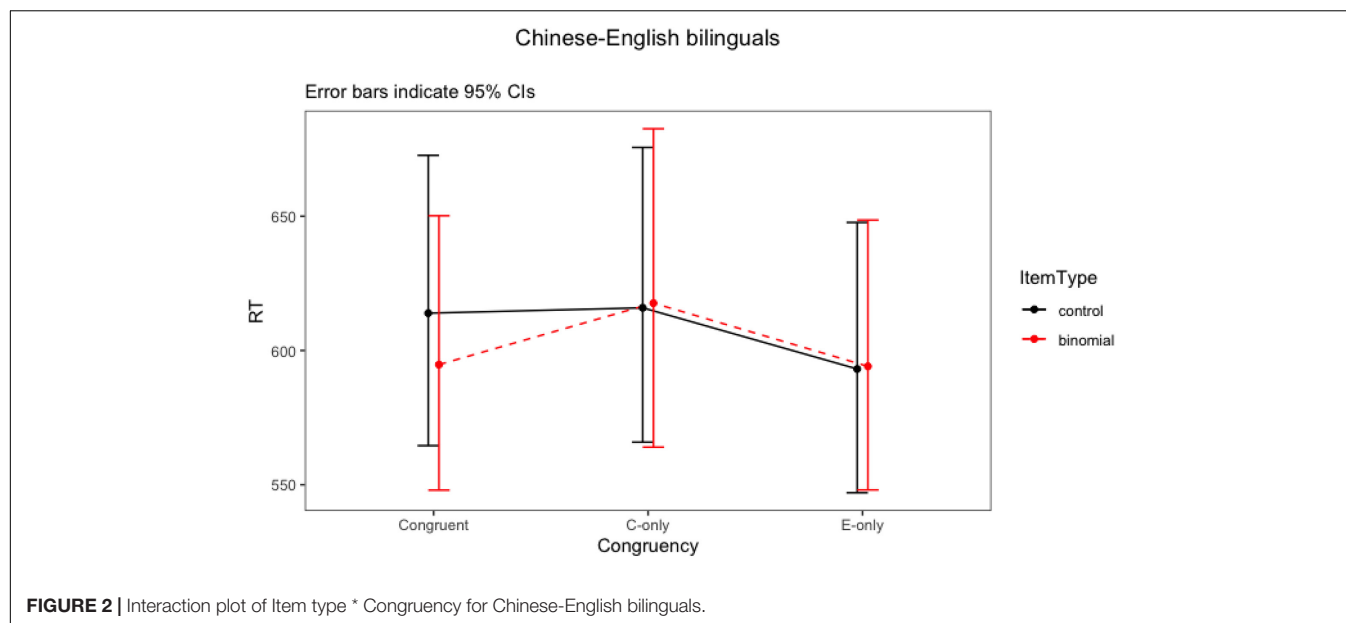


**FIGURE 2 |** Interaction plot of Item type \* Congruency for Chinese-English bilinguals.

and English-only binomials were matched in English phrase frequency ( $p = 0.94$ ) and did not show any difference in monolingual processing, we take the accelerated processing of congruent L2 binomials by Chinese-English bilinguals as evidence for the L1-L2 congruency effect. That is, CLI was observed in the processing of congruent L2 binomials by Chinese-English bilinguals.

### L1 MWE Activation Account

As expected, the English native speakers showed no priming for the translated Chinese-only binomials over controls, since these word sequences were unknown to these participants. Importantly, the Chinese-English bilinguals did not show a significant priming effect for the translated Chinese-only binomials either (cf. RQ3), suggesting that translated Chinese-only binomials were not processed as binomials in the L2. This finding is inconsistent with the L1 MWE activation account of L2 processing, as shown in previous studies on the processing of translated L1-only *idioms* (e.g., Carrol and Conklin, 2014, 2017; Carrol et al., 2016). Carrol and Conklin reported a processing advantage for translated Chinese-only idioms over matched controls with Chinese-English bilinguals, in a primed lexical decision task (Carrol and Conklin, 2014) and in an eye-tracking study (Carrol and Conklin, 2017). Moreover, in an eye-tracking study with Swedish-English bilinguals, Carrol et al. (2016) replicated and extended this finding; they showed that translated Swedish-only idioms showed the same degree of processing advantage as congruent idioms, and that there was no more facilitation for congruent idioms than for Swedish-only ones due to their additional experience in the L2. This led them to conclude that, over and above direct experience in the L2, L1 MWE knowledge directly affects how translation equivalents are processed in the L2. This discrepancy between the present results and those of Carrol et al. (2016) may be due to the type of MWEs (i.e., idioms vs. binomials) and to the

methodological differences between the studies, which will be further considered below.

Idioms are “strings of words whose figurative meaning does not necessarily derive from that of the constituent parts” (Cacciari, 2014, p. 267). That is, idioms have a figurative phrasal meaning and a literal meaning that reflects the meaning of their individual constituents. Thus, the processing advantage for idioms can come from form activation *and/or* from meaning activation (Carrol et al., 2016). Form activation refers to the recognition of specific word combinations presented in a particular order or configuration (i.e., lexical locus), while meaning activation refers to the understanding of the intended figurative phrasal meaning (i.e., conceptual locus) (Carrol et al., 2016). The robust advantage for translated L1-only idioms may come from meaning activation, although bilinguals may be unfamiliar with the form, when presented in the L2. For example, Beck and Weber (2016) found that translatable idioms (which have a matching concept *and* a word-for-word equivalent in L1) and untranslatable idioms (which have a matching concept *but* no translation equivalent in L1) produced comparable priming effect in proficient L1 German-L2 English bilinguals. This suggests that facilitation for the translated L1 idioms is likely to be driven by the conceptual overlap. In contrast, for literal MWEs (such as binomials), the processing advantage is likely

**TABLE 9 |** Results of *post hoc*, within-group tests of RTs for congruent, English-only, and Chinese-only items relative to the control items for Chinese-English bilinguals.

| Contrast    | Group | Estimate | Standard error | <i>t</i> | <i>p</i> | ED (msec) |
|-------------|-------|----------|----------------|----------|----------|-----------|
| Ctrl-Cngr   | CE    | 0.053    | 0.016          | 3.388    | 0.013    | 19        |
| Ctrl-E only | CE    | -0.003   | 0.015          | -0.187   | 0.99     | -1        |
| Ctrl-C only | CE    | -0.005   | 0.018          | -0.253   | 0.99     | -2        |

ED (estimated difference) is calculated with RTs to controls minus RTs to binomials.

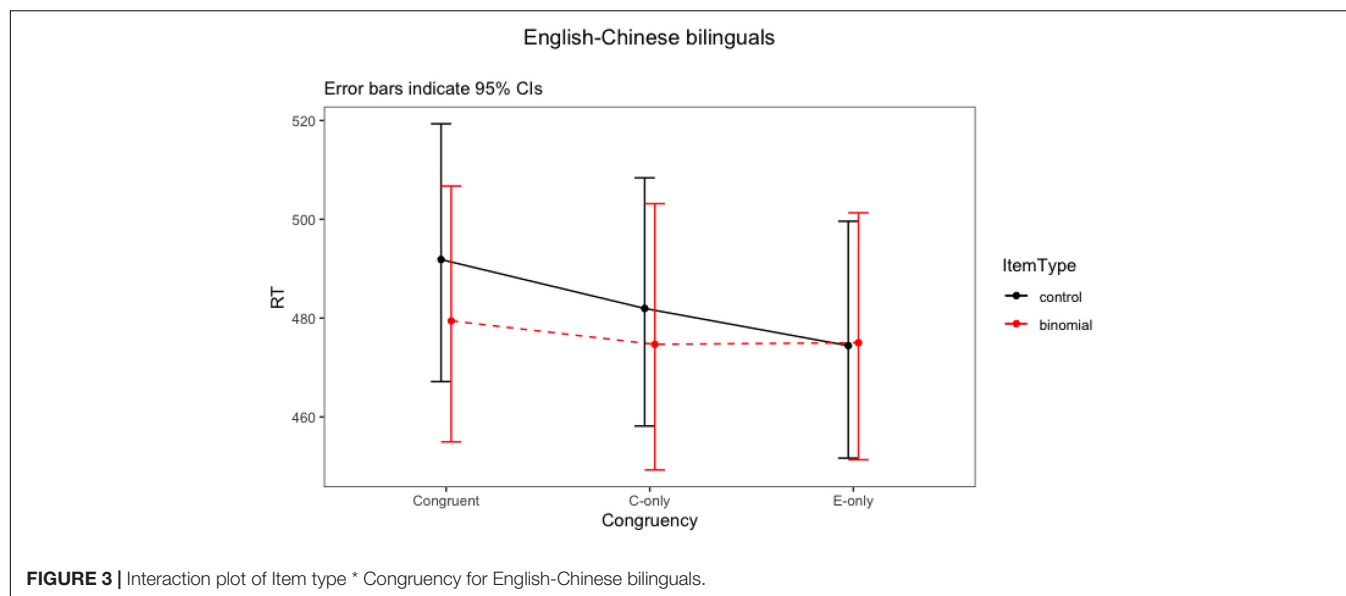


**TABLE 10 |** Results of *post hoc*, within-group tests of RTs for congruent, English-only, and Chinese-only items relative to the control items for English-Chinese bilinguals.

| Contrast    | Group | Estimate | Standard error | t      | p    | ED (msec) |
|-------------|-------|----------|----------------|--------|------|-----------|
| Ctrl-Cngr   | EC    | 0.053    | 0.019          | 2.717  | 0.11 | 12        |
| Ctrl-E only | EC    | -0.003   | 0.019          | -0.135 | 0.99 | -1        |
| Ctrl-C only | EC    | 0.032    | 0.027          | 1.19   | 0.99 | 7         |

ED (estimated difference) is calculated with RTs to controls minus RTs to binomials.

due to form activation, i.e., based on the cooccurrence of the component word forms in a particular order.

Methodologically, the difference between the present study and studies on idioms is that the latter examined the processing of idioms in sentence contexts (Carrol et al., 2016; Carrol and Conklin, 2017), while the present study looked at the processing of binomials out of context. A biasing context greatly increases predictability in the processing of idioms (Titone and Connine, 1999; Cieřlicka, 2013), which could have contributed to the translated L1-only idioms facilitation. Furthermore, Carrol and Conklin (2014, 2017) used very long idioms (e.g., *draw a snake and add ... feet*), which may have allowed participants to actively anticipate the completion to a phrase (Carrol et al., 2016). Critically, most of the studies that found facilitation for translated L1-only idioms employed eye-tracking, while the present study employed a lexical decision task. Speeded primed lexical decisions rely on lexical level activation processes that are mostly automatic. Therefore, we chose the primed lexical decision paradigm to test for automatic cross-language activation.

## L2 MWE Experience Account

Our findings are consistent with those reported in Wolter and Yamashita (2015, 2018). Wolter and Yamashita did not observe a processing advantage for translated Japanese-only collocations

compared to non-collocational matched controls with Japanese-English bilinguals in two response-based tasks: a double lexical decision task (Wolter and Yamashita, 2015) and an acceptability judgment task (Wolter and Yamashita, 2018). Both in the present study and in Wolter and Yamashita (2015, 2018), translated L1-only MWEs were processed as unknown word combinations, suggesting that there was no automatic activation of known L1 MWEs in L2 processing.

The absence of priming for translated L1-only MWEs is predicted by the L2 MWE experience account which, similar to usage- and exemplar-based acquisition and processing accounts, assumes that frequency of encounters with and use of a lexical item (words, MWEs) determines quality of its mental representations and its ease of processing (Langacker, 2000; Bybee, 2006). A plethora of empirical studies have shown that frequency plays a key role in MWE processing (e.g., Arnon and Snider, 2010; Siyanova-Chanturia et al., 2011). It is argued that due to their frequency, MWEs are processed faster than matched novel phrases by L1 as well as L2 speakers (for a review, see Siyanova-Chanturia and van Lancker Sidtis, 2019). Since translated L1-only MWEs do not exist in the participants' L2, they are unlikely to show a phrase frequency effect in the L2.

While the results for Chinese-English bilinguals suggest that known L1 MWEs are not automatically activated in the processing of the translated Chinese-only binomials, we are in no position to abandon the L1 MWE activation explanation entirely. For example, the original L1 MWEs may have been activated when the bilinguals read their L2 translation equivalents, but this activation may have been counteracted by the need to inhibit the non-target language (here, the participants' L1, Chinese), since the task was completed entirely in the L2 (Green, 1998). Additionally, the Chinese-English bilinguals' L1 may be inhibited, at the whole language level, in the context of their L2 immersion (Linck et al., 2009). Neurological studies have shown that competing information in the L1 needs to be suppressed to access information in an L2 (Abutalebi and Green,

2007; Pulido, 2021). Inhibiting L1 interference can improve L2 performance, in both immersion and non-immersion context (i.e., the L1 Regulation Hypothesis: Bogulski et al., 2019). In summary, the L1 inhibition necessitated by the experimental task and the country of residence contexts may have canceled out the possible activation of the L1 MWEs, resulting in no priming for translated Chinese-only binomials. In this case, the facilitation observed for the congruent L2 binomials could be due to their earlier acquisition by the bilinguals who, as a result, would have had a more extensive L2 processing experience with these binomials (as proposed in the L2 MWE experience account). This age of acquisition effect can also account for the congruency effect – the advantage in the processing of congruent over L2-only binomials. Since congruent and L2-only binomials were matched for L2 phrase frequency, there must be something other than L2 phrase frequency that contributed to the greater priming for congruent over L2-only MWEs. Age of acquisition of congruent L2 MWEs may well be such a factor. Wolter and colleagues (Wolter and Gyllstad, 2013; Wolter and Yamashita, 2015, 2018) argued that (1) congruent L2 MWEs are generally acquired earlier than incongruent L2-only MWEs, because acquisition is more straightforward when there is correspondence between the L1 and L2, and (2) earlier acquired congruent L2 MWEs are processed faster than later-acquired incongruent L2-only MWEs due to AoA effect. Because congruent MWEs share form (translation equivalents), structure (fixed word order) and referential meaning (same construct, e.g., *sun and moon*), they are more likely to be noticed in the L2 input and may be acquired faster (Yamashita and Jiang, 2010). One of potential mechanisms of the L1 transfer could be an initial strong declarative memory trace when encountering a congruent binomial in the L2 that exists in the learners' L1. This initial declarative knowledge can facilitate the gradual acquisition of procedural knowledge from input, thus, the multiword sequence is acquired procedurally and may be processed faster and more automatically than L2-only MWEs (Ullman, 2014). Wolter and Gyllstad (2011) found that congruent collocations were processed faster than incongruent L2-only collocations with Swedish-English bilinguals in a primed lexical decision task. They interpreted the finding as evidence for L1 influence on the *development* of L2 collocational knowledge. Similarly, Yamashita and Jiang (2010) found that Japanese-English bilinguals made fewer errors on congruent collocations than incongruent L2-only collocations in a phrase-acceptability judgment task, irrespective of their L2 proficiency. This suggests that congruent L2 collocations show an acquisition advantage at the early stages of L2 learning. Incongruent L2-only MWEs, on the other hand, may need more repeated exposure to the L2 to be acquired. This account could also explain why no priming was observed for incongruent English-only binomials over the controls for the Chinese-English bilinguals in the present study. Similarly, in a lexical decision task, Wolter and Yamashita (2015) found that Japanese-English bilinguals did not produce accelerated processing for L2-only collocations either. It is thus plausible that the processing advantage for congruent over L2-only formulaic sequences is due to their age of acquisition. However, further empirical support is needed for

the proposition that congruent MWEs are better noticed in the L2 input and are acquired earlier than incongruent L2-only MWEs (e.g., Arnon et al., 2017).

## Cross-Language Influences From L2 to L1

With respect to the performance of the English-Chinese bilinguals, our key findings were as follows. Unlike the English monolingual participants, the English-Chinese bilinguals did not show significant facilitation in the processing of the final word in the *English-only* binomial phrases (*bread and butter*) compared to the control phrases (*toast and butter*), but they showed a clear trend toward priming for the *congruent* binomials (*sun and moon*) compared to control phrases (*star and moon*). They processed congruent binomials quantitatively faster than their controls (mean difference = 20 ms, model estimate = 12 ms). However, after applying a correction for multiple comparison, the priming did not reach statistical significance ( $p = 0.11$ ). Thus, compared with the Chinese-English bilinguals who showed significant priming in the processing of congruent binomials and a clear congruency advantage over English-only binomials, English-Chinese bilinguals showed only a weak congruency advantage (cf. RQ2). Finally, similar to the English monolinguals and Chinese-English bilinguals, the English-Chinese bilinguals showed no processing advantage for the translated Chinese-only binomials (*wisdom and strength*) over controls (cf. RQ3). We discuss each of these findings below.

### The Inhibition of an L1

The finding that the English-Chinese bilinguals showed no processing advantage for English-only binomials over controls seems inconsistent with the literature on MWE processing in L1 speakers. It has been established that L1 speakers can recognize, read and respond to MWEs significantly faster than matched novel strings of language (Arnon and Snider, 2010; Durrant and Doherty, 2010; Vilkaite, 2016; Siyanova-Chanturia et al., 2017). In fact, we also observed a significant priming effect for English-only binomials for the English monolingual controls. What, then, might have caused the absence of priming for English-only binomials for the English-Chinese bilingual speakers, who performed the task in their native and dominant language?

One possibility is that the L1 of the English-Chinese bilinguals had been strongly inhibited in the L2 immersion environment (while studying Chinese in China). When they had to switch back to their strongly inhibited L1, for the purpose of completing the experiment, their L1 processing could have been impaired. The result that the mean RTs on L1 (English) lexical decisions were somewhat slower for the English-Chinese bilinguals than for the English monolinguals (512 ms vs. 497 ms;  $p < 0.0001$ ) provides some evidence to support this conjecture. It has been shown that, after immersion in a foreign language, even just for a few months, bilinguals may experience delay when retrieving L1 words (Linck et al., 2009; Baus et al., 2013). Immersion is argued to enable bilinguals to attenuate the activity of the L1, thus better controlling L1 lexical competition and facilitating L2 learning (Linck et al., 2009). For instance, in a comprehension

task (translation recognition), Linck et al. (2009) found that the immersed English-Spanish bilinguals showed no sensitivity to English distractors which had form overlap with the presented Spanish words (e.g., *cara-card*). The results were interpreted as evidence that immersed bilinguals suppress the visually presented distractors from intruding on their judgments, and that L1 was inhibited frequently during immersion to facilitate L2 learning. Recent evidence from classroom learning also indicates that the inhibition of L1 equivalents improves learning and retrieval of L2 MWEs in an L1-speaking environment (Pulido and Dussias, 2020; Pulido, 2021).

According to the Inhibitory Control model (Green, 1998), the non-target language is inhibited, preventing it from disrupting the selection of target language words. The amount of inhibition applied to the non-target language is proportional to the baseline strength of its activation. The more dominant the language, the stronger inhibition is needed. Since the L1 of an unbalanced bilingual is dominant, it is strongly suppressed whenever bilinguals need to use L2. As a result, the cost of reactivating L1 after using L2 is likely to be greater than a switch in the opposite direction (Mosca and de Bot, 2017; Wodniecka et al., 2020), having a greater effect on L1 performance. Numerous studies have shown that switching costs are larger for the stronger than for the weaker language (i.e., asymmetrical switching costs) (Meuter and Allport, 1999; Jackson et al., 2001; Macizo et al., 2012). For our unbalanced English-Chinese bilinguals, the L1 had to be strongly inhibited to enable them to use L2 in the immersion context. Switching back to their strongly suppressed L1, in order to perform an L1 lexical decision task, likely came at a cost. The absence of priming for English-only binomials in the English-Chinese group may have been a result of their weakened L1 performance.

The Chinese-English bilinguals who reported longer years of L2 exposure (3.9 vs. 1.9) and high proficiency in English may have been more balanced than the English-Chinese bilinguals and, therefore, may not have needed to inhibit their L1 as strongly. This would explain why their L1 could have been more readily activated during the processing of the congruent L2 binomials. This account is compatible with the extended Inhibitory Control model that is based on the language balance model (Wodniecka et al., 2020), which holds that the amount of inhibition applied to L1 during L2 use is related to the relative balance between the two languages. Studies have shown that when the two languages of a bilingual speaker are relatively balanced, the switching costs between languages becomes comparable, i.e., symmetrical switching costs (Christoffels et al., 2007; Schwieter and Sunderman, 2008; Declerck et al., 2013). In order to test this account, a follow-up study would need to compare the processing of L1 MWEs by bilinguals in an L2 immersion context and bilinguals in their L1 context. This would allow us to examine whether the impact of immersion on the L1 MWE processing is similar to that reported for single words (Linck et al., 2009; Morales et al., 2014).

## L2 Influence on L1 MWE Processing

This difficulty of retrieving the dominant L1 in the L2 immersion context may also explain our findings for the processing of

congruent L1 binomials by the English-Chinese bilinguals. The English-Chinese bilinguals showed some facilitation in the processing of the final word in congruent L1 binomials relative to control phrases, although the result was less robust than that observed for the Chinese-English bilinguals. This result is compatible with their performance on English-only binomials, suggesting attenuated L1 access due to L1 inhibition. We observed a clear trend toward priming for congruent L1 binomials that suggests possible activation of known corresponding L2 binomials. Since English-only and congruent L1 binomials were matched in L1 phrase frequency and our English monolingual controls showed comparable facilitation for both, activation of L2 binomial equivalents during the L1 task by the English-Chinese bilinguals seems to be the likely explanation of this priming trend for congruent L1 (but not English-only) binomials. This is evidence of cross-language influence in the L2-L1 direction in an entirely within-L1 task. This result is not unlike the findings of an automatic activation of single words in the weaker language in mixed stimulus lists (Dijkstra et al., 2000) and in L1-only lists (e.g., van Hell and Dijkstra, 2002). Our finding suggests that known L2 MWEs may be automatically activated in L1 processing, leading to the faster processing of MWEs that exist in both languages.

Finally, English-Chinese bilinguals showed no facilitation for translated Chinese-only binomials over controls. The same pattern of results was observed in English monolinguals and Chinese-English bilinguals. This indicates that there was no activation of translated Chinese-only MWEs (i.e., L2-only in the case of English-Chinese bilinguals). It is not surprising given that there was no activation for translated Chinese-only binomials over controls in Chinese-English bilinguals. In other words, the effects in the L2-L1 direction were less likely to take place when no such effects were observed in the L1-L2 direction, because CLI in the L1-L2 direction is normally stronger than in the opposite direction.

## CONCLUSION

In this study, we employed binomial expressions in order to examine crosslinguistic influences in the processing of MWEs in both directions. The results from Chinese-English bilinguals show that congruent L2 binomials showed greater priming effects than English-only binomials and that translated Chinese-only binomials showed no priming effect. We take these results as evidence that L1 influences the processing of binomials in the L2 and interpret them as supporting the L2 MWE experience account, according to which congruent MWEs should be processed faster than incongruent MWEs because they are noticed and acquired earlier due to the positive L1 transfer. English-Chinese bilinguals showed no priming for English-only binomials, but a clear priming trend for congruent binomials. These results support the view that L1 may be inhibited in L2 learning and immersion contexts and, thus, switching back to L1 may come at a cost. The results also support the view that crosslinguistic influence can occur from the non-dominant L2 to the dominant L1, even in an entirely within-L1 task. Thus,

we conclude that crosslinguistic influences in the processing of binomials are bidirectional, although the influence in the direction of L1-L2 is stronger than in the reverse direction of L2-L1. This conclusion is in line with studies with bilingual word processing which suggest that crosslinguistic influences are bi-directional. The present study is the first study, to our knowledge, that investigated bi-directional cross-language influences in the processing of binomials – a less commonly studied type of MWEs.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Victoria University of Wellington, New Zealand. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LD, IE, and AS-C contributed to conception and design of the study. LD carried out the experiment and collected the data. LD and IE conducted the statistical analysis. LD wrote the first draft

of the manuscript. IE and AS-C wrote sections of the manuscript. All the authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This research was supported by the Innovation Funds Plan of Henan University of Technology (2021-SKCXTD-13) to the research group of Foreign Language, Literature and International Sinology, and by Wellington Doctoral Submission scholarship and the Faculty of Humanities and Social Sciences Research Grant #220046 awarded to LD.

## ACKNOWLEDGMENTS

We are grateful to the two reviewers, whose comments and suggestions helped to make this a better article.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.666520/full#supplementary-material>

**Supplementary Appendix 1** | Binomials and control items used in the experiment.

**Supplementary Appendix 2** | Filler and non-word items used in the experiment.

## REFERENCES

- Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003
- Arnon, I., McCauley, S. M., and Christiansen, M. H. (2017). Digging up the building blocks of language: age-of-acquisition effects for multiword phrases. *J. Mem. Lang.* 92, 265–280. doi: 10.1016/j.jml.2016.07.004
- Arnon, I., and Snider, N. (2010). More than words: frequency effects for multi-word phrases. *J. Mem. Lang.* 62, 67–82. doi: 10.1016/j.jml.2009.09.005
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv arXiv:1506.04967*.
- Baus, C., Costa, A., and Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychol.* 142, 402–409. doi: 10.1016/j.actpsy.2013.01.010
- Beck, S. D., and Weber, A. (2016). Bilingual and monolingual idiom processing is cut from the same cloth: the role of the L1 in literal and figurative meaning activation. *Front. Psychol.* 7:1350. doi: 10.3389/fpsyg.2016.01350
- Benor, S. B., and Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82, 233–278. doi: 10.1353/lan.2006.0077
- Bogulski, C. A., Bice, K., and Kroll, J. F. (2019). Bilingualism as a desirable difficulty: advantages in word learning depend on regulation of the dominant language. *Bilingual. Lang. Cogn.* 22, 1052–1067. doi: 10.1017/s1366728918000858
- Brysbaert, M., and Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *J. Cogn.* 1, 1–20.
- Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language* 82, 711–733. doi: 10.1353/lan.2006.0186
- Cacciari, C. (2014). Processing multiword idiomatic strings: many words in one? *Ment. Lexic.* 9, 267–293. doi: 10.1075/ml.9.2.05cac
- Carrol, G., and Conklin, K. (2014). Getting your wires crossed: evidence for fast processing of L1 idioms in an L2. *Bilingual. Lang. Cogn.* 17, 784–797. doi: 10.1017/s1366728913000795
- Carrol, G., and Conklin, K. (2017). Cross language lexical priming extends to formulaic units: evidence from eye-tracking suggests that this idea 'has legs'. *Bilingual. Lang. Cogn.* 20, 299–317. doi: 10.1017/s1366728915000103
- Carrol, G., and Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Lang. Speech* 63, 95–122. doi: 10.1177/0023830918823230
- Carrol, G., Conklin, K., and Gyllstad, H. (2016). Found in translation: the influence of the L1 on the reading of idioms in a L2. *Stud. Sec. Lang. Acqu.* 38, 403–443. doi: 10.1017/s0272263115000492
- Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208. doi: 10.1016/j.brainres.2007.01.137
- Cieslicka, A. B. (2013). Do nonnative language speakers chew the fat and spill the beans with different brain hemispheres? Investigating idiom decomposability with the divided visual field paradigm. *J. Psycholinguist. Res.* 42, 475–503. doi: 10.1007/s10936-012-9232-4
- Conklin, K. (2020). "Processing single-word and multiword items," in *The Routledge handbook of vocabulary studies*, ed. W. Stuart (New York: Routledge), 174–188.
- Conklin, K., and Carrol, G. (2018). "First language influence on the processing of formulaic language in a second language," in *Understanding formulaic language: A second language acquisition perspective*, eds A. Siyanova-Chanturia and A. Pellicer-Sánchez (New York, NY: Routledge), 62–77. doi: 10.4324/9781315206615-4
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 Million Words, 1990-Present*. Denton, TX: University of North Texas Libraries.

- Declerck, M., Philipp, A. M., and Koch, I. (2013). Bilingual control: sequential memory in language switching. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1793–1806. doi: 10.1037/a0033094
- Degani, T., Prior, A., and Tokowicz, N. (2011). Bidirectional transfer: the effect of sharing a translation. *J. Cogn. Psychol.* 23, 18–28. doi: 10.1080/20445911.2011.445986
- Dijkstra, T., Timmermans, M., and Schriefers, H. (2000). On being blinded by your other language: effects of task demands on interlingual homograph recognition. *J. Mem. Lang.* 42, 445–464. doi: 10.1006/jmla.1999.2697
- Durrant, P., and Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corp. Linguist. Linguist. Theory* 6, 125–155.
- Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: evidence for nonselective phonological activation in bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1340–1359. doi: 10.1037/0278-7393.31.6.1340
- Ellis, A. W., and Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 515–523. doi: 10.1037/0278-7393.24.2.515
- Finkbeiner, M., Forster, K., Nicol, J., and Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *J. Mem. Lang.* 51, 1–22. doi: 10.1016/j.jml.2004.01.004
- Forster, K. I., and Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/bf03195503
- Gollan, T. H., Forster, K. I., and Frost, R. (1997). Translation priming with different scripts: masked priming with cognates and noncognates in Hebrew–English bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1122–1139. doi: 10.1037/0278-7393.23.5.1122
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingual. Lang. Cogn.* 1, 67–81. doi: 10.1017/s1366728998000133
- Hernández, M., Costa, A., and Arnon, I. (2016). More than words: multiword frequency effects in non-native speakers. *Lang. Cogn. Neurosci.* 31, 785–800. doi: 10.1080/23273798.2016.1152389
- Jackson, G. M., Swainson, R., Cunningham, R., and Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition* 4, 169–178. doi: 10.1017/s1366728901000268
- Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingual. Lang. Cogn.* 2, 59–75. doi: 10.1017/s1366728999000152
- Jiang, N., and Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Mod. Lang. J.* 91, 433–445. doi: 10.1111/j.1540-4781.2007.00589.x
- Juhász, B. J., and Rayner, K. (2006). The role of age of acquisition and word frequency in reading: evidence from eye fixation durations. *Vis. Cogn.* 13, 846–863. doi: 10.1080/13506280544000075
- Keatley, C. W., Spinks, J. A., and De Gelder, B. (1994). Asymmetrical cross-language priming effects. *Mem. Cogn.* 22, 70–84. doi: 10.3758/bf03202763
- Kim, J., and Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *J. Mem. Lang.* 49, 484–499. doi: 10.1016/s0749-596x(03)00093-7
- Kroll, J. F., and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *J. Cogn. Psychol.* 25, 497–514. doi: 10.1080/20445911.2013.799170
- Kroll, J. F., and De Groot, A. M. (2009). *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). Package ‘lmerTest’. R package version, 2(0).
- Langacker, R. (2000). “A dynamic usage-based model,” in *Usage based models of language*, eds S. Kemmer and M. Barlow (Stanford: CSLI publications, Center for the Study of Language and Information), 1–64.
- Larson, M. J., Clayson, P. E., and Clawson, A. (2014). Making sense of all the conflict: a theoretical review and critique of conflict-related ERPs. *Int. J. Psychophysiol.* 93, 283–297. doi: 10.1016/j.ijpsycho.2014.06.007
- Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.4.1.
- Li, K., and Ho, Y. (2016). A corpus-based comparative study on George Orwell’s 1984 Chinese translation strategies. *Stud. Literat. Lang.* 13, 26–33.
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing access to the native language while immersed in a second language: evidence for the role of inhibition in second-language learning. *Psychol. Sci.* 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x
- Macizo, P., Bajo, T., and Paolieri, D. (2012). Language switching and language competition. *Sec. Lang. Res.* 28, 131–149. doi: 10.1177/0267658311434893
- Matsuno, K. (2017). Processing collocations: do native speakers and second language learners simultaneously access prefabricated patterns and each single word? *J. Eur. Sec. Lang. Assoc.* 1, 61–72. doi: 10.22599/jesla.17
- McEnery, A., and Xiao, Z. (2004). The Lancaster corpus of mandarin Chinese: a corpus for monolingual and contrastive language study. *Religion* 17, 3–4.
- McNamara, T. P. (2005). *Semantic Priming: Perspectives From Memory and Word Recognition*. Hove: Psychology Press.
- Meuter, R. F., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602
- Millar, N. (2011). The processing of malformed formulaic language. *Appl. Linguist.* 32, 129–148. doi: 10.1093/applin/amq035
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Oxford University Press.
- Morales, L., Paolieri, D., Cubelli, R., and Bajo, M. T. (2014). Transfer of Spanish grammatical gender to English: evidence from immersed and non-immersed bilinguals. *Bilingualism: Language and Cognition* 17, 700–708. doi: 10.1017/s1366728914000017
- Morrison, C. M., and Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 116–133. doi: 10.1037/0278-7393.21.1.116
- Mosca, M., and de Bot, K. (2017). Bilingual language switching: production vs. recognition. *Front. Psychol.* 8:934. doi: 10.3389/fpsyg.2017.00934
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*, Vol. 14. Amsterdam: John Benjamins.
- Pulido, M. F. (2021). Native language inhibition predicts more successful second language learning: evidence of two ERP pathways during learning. *Neuropsychologia* 152:107732. doi: 10.1016/j.neuropsychologia.2020.107732
- Pulido, M. F., and Dussias, P. E. (2020). Desirable difficulties while learning collocations in a second language: conditions that induce L1 interference improve learning. *Bilingual. Lang. Cogn.* 23, 652–667. doi: 10.1017/s1366728919000622
- R Core Team (2016). *R Core Team*. Vienna: R Foundation for Statistical Computing.
- Rastle, K., Harrington, J., and Coltheart, M. (2002). 358,534 nonwords: the ARC nonword database. *Q. J. Exp. Psychol. Sect. A* 55, 1339–1362. doi: 10.1080/02724980244000099
- Römer, U., O’Donnell, M. B., and Ellis, N. C. (2014). Second language learner knowledge of verb–argument constructions: effects of language transfer and typology. *Mod. Lang. J.* 98, 952–975. doi: 10.1111/modl.12149
- Schoonbaert, S., Duyck, W., Brysbaert, M., and Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: making sense of the findings. *Mem. Cogn.* 37, 569–586. doi: 10.3758/mc.37.5.569
- Schoonbaert, S., Hartsuiker, R. J., and Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: evidence from syntactic priming. *J. Mem. Lang.* 56, 153–171. doi: 10.1016/j.jml.2006.10.002
- Schwieter, J. W., and Sunderman, G. (2008). Language switching in bilingual speech production: in search of the language-specific selection mechanism. *Ment. Lexic.* 3, 214–238. doi: 10.1075/ml.3.2.06sch
- Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., and van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain Lang.* 175, 111–122. doi: 10.1016/j.bandl.2017.10.004
- Siyanova-Chanturia, A., Conklin, K., and van Heuven, W. J. (2011). Seeing a phrase “time and again” matters: the role of phrasal frequency in the processing of multiword sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 776–784. doi: 10.1037/a0022531
- Siyanova-Chanturia, A., and Martinez, R. (2015). The idiom principle revisited. *Appl. Linguist.* 36, 549–569.

- Siyanova-Chanturia, A., and Van Lancker Sittis, D. (2019). "What online processing tells us about formulaic language," in *Understanding formulaic language: A second language acquisition perspective*, eds A. Siyanova-Chanturia and A. Pellicer-Sánchez (New York, NY: Routledge), 38–61. doi: 10.4324/9781315206615-3
- Sonbul, S., El-Dakhs, D. A. S., and Al-Otaibi, H. (2020). Productive versus receptive L2 knowledge of polysemous phrasal verbs: a comparison of determining factors. *System* 95:102361. doi: 10.1016/j.system.2020.102361
- Sprenger, S. A., Levelt, W. J., and Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *J. Mem. Lang.* 54, 161–184. doi: 10.1016/j.jml.2005.11.001
- Titone, D., Columbus, G., Whitford, V., Mercier, J., and Libben, M. (2015). "Contrasting bilingual and monolingual idiom processing" in *Bilingual figurative language processing*, eds R. R. Heredia and A. B. Cieslicka (Cambridge: Cambridge University Press), 171–207. doi: 10.1017/cbo9781139342100.011
- Titone, D., and Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *J. Pragmat.* 31, 1655–1674. doi: 10.1016/s0378-2166(99)00008-9
- Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks. *Lang. Learn.* 61, 569–613. doi: 10.1111/j.1467-9922.2010.00622.x
- Ullman, M. T. (2014). "The declarative/procedural model: a neurobiologically motivated theory of first and second language," in *Theories in second language acquisition: An introduction*, eds B. Vanpatten and J. Williams (New York, NY: Routledge), 135–160.
- van Hell, J. G., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonom. Bull. Rev.* 9, 780–789. doi: 10.3758/bf03196335
- Vilkaite, L. (2016). Are nonadjacent collocations processed faster? *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 1632–1642. doi: 10.1037/xlm0000259
- Whitford, V., and Titone, D. (2019). Lexical entrenchment and cross-language activation: two sides of the same coin for bilingual reading across the adult lifespan. *Bilingual. Lang. Cogn.* 22, 58–77. doi: 10.1017/s1366728917000554
- Wodniecka, Z., Casado, A., Kalamala, P., Marecka, M., Timmer, K., and Wolna, A. (2020). "The dynamics of language experience and how it affects language and cognition," in *Adult and second language learning*, eds K. D. Federmeier and H.-W. Huang (Cambridge: Academic Press).
- Wolter, B., and Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Appl. Linguist.* 32, 430–449. doi: 10.1093/applin/amr011
- Wolter, B., and Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Stud. Sec. Lang. Acqu.* 35, 451–482. doi: 10.1017/s0272263113000107
- Wolter, B., and Yamashita, J. (2015). Processing collocations in a second language: a case of first language activation? *Appl. Psycholinguist.* 36, 1193–1221. doi: 10.1017/s0142716414000113
- Wolter, B., and Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing. *Stud. Sec. Lang. Acqu.* 40, 395–416. doi: 10.1017/s0272263117000237
- Wu, Y. J., Cristino, F., Leek, C., and Thierry, G. (2013). Non-selective lexical access in bilinguals is spontaneous and independent of input monitoring: evidence from eye tracking. *Cognition* 129, 418–425. doi: 10.1016/j.cognition.2013.08.005
- Wu, Y. J., and Thierry, G. (2010). Chinese–English bilinguals reading English hear Chinese. *J. Neurosci.* 30, 7646–7651. doi: 10.1523/jneurosci.1602-10.2010
- Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *J. Sec. Lang. Stud.* 1, 60–78. doi: 10.1075/jsls.17024.yam
- Yamashita, J., and Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Q.* 44, 647–668. doi: 10.5054/tq.2010.235998
- Zeng, T., Branigan, H. P., and Pickering, M. J. (2020). Do bilinguals represent between-language relationships beyond the word level in their lexicon? *J. Neurolinguist.* 55:100892. doi: 10.1016/j.jneuroling.2020.100892
- Zhan, W., Guo, R., and Chen, Y. (2003). *The CCL Corpus of Chinese Texts: 700 Million Chinese Characters, the 11th Century B.C.-present*. Available online at: <http://ccl.pku.edu.cn/8080> (accessed May, 2018).
- Zhang, T., Van Heuven, W. J., and Conklin, K. (2011). Fast automatic translation and morphological decomposition in Chinese-English bilinguals. *Psychol. Sci.* 22, 1237–1242. doi: 10.1177/0956797611421492

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Du, Elgort and Siyanova-Chanturia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Hands Down: Cognate Effects Persist During Written Word Production

Evvy Woumans<sup>1\*</sup>, Robin Clauws<sup>2</sup> and Wouter Duyck<sup>2</sup>

<sup>1</sup> Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium, <sup>2</sup> Department of Experimental Psychology, Ghent University, Ghent, Belgium

## OPEN ACCESS

### Edited by:

Zofia Wodniecka,  
Jagiellonian University, Poland

### Reviewed by:

Sara Guediche,  
Basque Center on Cognition, Brain  
and Language, Spain  
Pilar Ferré Romeu,  
University of Rovira i Virgili, Spain  
Eugenia Navarra,  
Basque Center on Cognition, Brain  
and Language, San Sebastian, Spain,  
in collaboration with reviewer SG.

### \*Correspondence:

Evvy Woumans  
evvy.woumans@ugent.be

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 29 December 2020

**Accepted:** 07 June 2021

**Published:** 05 July 2021

### Citation:

Woumans E, Clauws R and  
Duyck W (2021) Hands Down:  
Cognate Effects Persist During  
Written Word Production.  
Front. Psychol. 12:647362.  
doi: 10.3389/fpsyg.2021.647362

Words that share form and meaning across two or more languages (i.e., cognates) are generally processed faster than control words (non-cognates) by bilinguals speaking these languages. This so-called cognate effect is considered to be a demonstration of language non-selectivity during bilingual lexical access. Still, research up till now has focused mainly on visual and auditory comprehension. For production, research is almost exclusively limited to speech, leaving written production out of the equation. Hence, the goal of the current study was to examine whether bilinguals activate representations from both languages during typewriting. Dutch-English bilinguals completed second-language written sentences with names of displayed pictures. Low-constraint sentences yielded a cognate facilitation effect, whereas high-constraint sentences did not. These findings suggest that co-activation of similar words across languages also occurs during written production, just as in reading and speaking. Also, the interaction effect with sentence constraint shows that grammatical and semantic sentence restrictions may overrule interlingual facilitation effects.

**Keywords:** cognate effect, typewriting, picture naming, sentence context, word production, language non-selective activation, bilingualism

## INTRODUCTION

If a bilingual reads, hears, or produces a word, do they activate a representation of that word in one or multiple languages? Much research has sought to answer this question and different theories have been put to the test. At present, there is general consensus on an integrated lexicon (see Brysbaert and Duyck, 2010) or a segregated lexicon that is activated in parallel (Dijkstra et al., 2019). Evidence to support this idea comes from studies employing interlingual homonyms or cognates to demonstrate language non-selective activation (e.g., Dijkstra et al., 1999; Costa et al., 2000; Lagrou et al., 2011a,b). Interlingual homonyms are words that share spelling but not meaning within or across languages, such as *boot* in English (meaning footwear) and *boot* in Dutch (meaning a vessel). We may distinguish between homographs (words that are typographically similar) and homophones (words that are phonologically similar). In contrast, cognates are words that share form (completely or predominantly) and meaning across languages, such as the Dutch and English word *film*. The cognate facilitation effect states that cognates are read (Cop et al., 2017), heard (Blumenfeld and Marian, 2007), and spoken (Costa et al., 2000) faster than non-cognates by bilinguals speaking those specific languages, whereas cross-linguistic homophones seem to interfere with bilingual language processing, in listening (Lagrou et al., 2015) and reading

(Dijkstra et al., 1999), just like homographs (see Dijkstra et al., 2000, for reading; Lagrou et al., 2011a,b; for listening; Jared and Szucs, 2002, for speaking).

Findings such as these show that words of each language can be activated in both first (L1) and second (L2) language processing and support both an integrated lexicon and language non-selective access. Still, looking at the body of established research on bilingual lexical access, it is clear that lot of research has mainly focused on word recognition (Caramazza and Brones, 1979; Dijkstra et al., 1999; Spivey and Marian, 1999; Schwartz et al., 2007; Lagrou et al., 2011a), and to a lesser extent on word production (Costa et al., 2000; Hoshino and Kroll, 2008). Studies on cross-lingual activation during written word production, however, are virtually non-existent. Hence, it was the current study's aim to fill in the gap for written word production. In addition, studies on word recognition and speaking mostly investigated isolated word processing, although the grammatical and semantic restrictions that arise from sentences may be important modulators of cross-lingual lexical activation. It is only later that sentence studies were carried out in these modalities (e.g., Duyck et al., 2007; Van Assche et al., 2009). Here, as a second aim, we will also investigate the impact of sentence constraint for written word production, and more specifically for typewriting.

## Examining Lexical Access Using Isolated Word Paradigms

Research into bilingual lexical access has often employed lexical decision paradigms to obtain more insight into bilingual language processing. A pioneering study by Caramazza and Brones (1979) had Spanish-English subjects classify strings of letters as either English or Spanish words or non-words. Also included were cognates, both in the blocked (L1 or L2 only) and mixed (L1 and L2) conditions. A cognate facilitation effect was found in the L2 blocked condition and in the mixed condition. Since then, this facilitation effect has been replicated by a number of studies (e.g., Dijkstra et al., 1999; Schwartz et al., 2007). Most notably, facilitatory effects of cross-linguistic overlap also appear to be progressive. In a study among trilingual subjects, Lemhöfer et al. (2004) employed both words that were cognates across two languages as well as words that were cognates across three languages. When performing a lexical decision task in their third language German, the Dutch-English-German trilinguals responded faster for so-called “double cognates” (with overlap in Dutch and German) as opposed to control words, but even faster responses were produced for triple cognates (with overlap in Dutch, German, and English). These findings again supported the view of language non-selective access, and furthermore implied that all languages known to an individual may affect word recognition and activation.

In a more elaborate study not only constricted to cognates but also containing homonyms, Dijkstra et al. (1999) explored lexical decision performance on English words that varied according to the degree with which they shared orthography, phonology, semantics, or some combination of the three codes with Dutch words. Dutch-English bilinguals showed faster response latencies for words that shared orthography (i.e.,

homographs) or a combination of orthography and semantics (i.e., cognates), supporting the language non-selective access hypothesis. In contrast, recognition latencies were delayed when words shared only phonology (i.e., homophones), which the authors explained as an inhibitory effect. A given letter string may activate all compatible phonological codes independent of language (e.g., Brysbaert et al., 1999), but the activated non-identical phonological lexical representations may compete at a lexical level, resulting in a delayed identification of the item in the target language (i.e., lateral inhibition).

Important to note, however, is that interlingual homographs and even cognates may also serve as inhibitors of lexical access when presented under constricting circumstances. When Dutch-English bilinguals were asked to respond to interlingual homographs in an English lexical decision task with only English words in the stimulus list, latencies for homographs and control words did not differ. However, when half of the non-words were replaced by Dutch words requiring a no-response in this task, latencies for homographs slowed down substantially (Dijkstra et al., 1998). Similarly, Dutch-English bilinguals demonstrated a cognate inhibition effect when performing an English lexical decision task where the non-words were replaced by Dutch words (Vanlangendonck et al., 2020). This adaptation in context (purely English versus English mixed with Dutch) completely reversed the cognate facilitation effect. In contrast, cognate facilitation does seem to uphold in mixed conditions where stimuli in both languages require a yes-response (i.e., generalized lexical decision) (Lemhöfer and Dijkstra, 2004).

Although the abovementioned studies were restricted to visual word recognition, evidence suggests that also for spoken word recognition, lexical access is non-selective. For instance, a study by Lagrou et al. (2011a) provided confirmation for the findings of Dijkstra et al. (1999), when they demonstrated a similar delay in response latencies when bilinguals had to respond to homophones in an auditory lexical decision task. Furthermore, an early study by Spivey and Marian (1999) hypothesized that bilinguals might be distracted by words from their one language when doing a task in their other language, if the initial phonemes of the words overlap. This hypothesis was based on previous findings from monolingual research, which suggest that when subjects produce a spoken word, all words starting with the same sounds are initially activated (e.g., Marslen-Wilson, 1987). Employing a visual word paradigm, Spivey and Marian presented Russian-English bilinguals with verbal instructions, which told them to pick up a stamp (e.g., *marku* in Russian) while one of the distractor objects was a *marker* (a word starting with the same phonemes). Confirming the author's hypothesis, the subjects looked more toward the marker than any of the other two objects. Similar results were found when English was the language of instruction.

Although a study by Weber and Cutler (2004) was able to replicate the cross-language finding of Spivey and Marian (1999) when instructions were given in L2, this was not the case for L1. The authors explained the difference by stating that Spivey and Marian had tested their participants in an L2 environment (Russian students studying at an American university), whereas Weber and Cutler studied their participants

in an L1 environment. However, Lagrou et al. (2013) rightly noted that the phoneme overlap between the targets and the distractors in the study by Weber and Cutler was very small (i.e., often only one phoneme). When these authors repeated the experiment with stimuli that elicited more cross-lingual activation (i.e., words with more overlap), they were able to confirm cross-language activation for spoken word recognition also in an L1 context.

Whereas the studies described above focused on word recognition, there is also evidence that confirms non-selective access in word production. Costa et al. (2000), for instance, found that Catalan-Spanish bilinguals were faster at naming cognate words in a picture naming task, which they performed in their L2 Spanish. Although Catalan and Spanish share many linguistic features, the finding has also been replicated among different-script bilinguals. In a study by Hoshino and Kroll (2008), Japanese-English bilinguals named cognate and non-cognate pictures in their L2 English and they also demonstrated a cognate facilitation effect. The authors concluded that this outcome implies there is cross-language activation of phonology even for different-script bilinguals.

## Does Sentence Constraint Alter Bilingual Lexical Access?

Until now, we have primarily focused on studies reporting cognate and interlingual homograph effects on words presented in isolation, as this was the prime interest of literature. However, the outcome of these studies may not be representative for natural language processing, as language context (e.g., through text and situation) usually provides bilinguals with a clear cue for which language requires activation, or even which word. Looking at sentence context, one may argue that if an unfolding sentence is predictive of an upcoming word, cognate effects for this particular word may disappear. Indeed, previous work shows that when sentences are low semantically constraining and therefore not predictive, cognate effects still appear, whereas this is not the case for highly constraining and hence predictive sentences (e.g., Schwartz and Kroll, 2006; Van Hell and De Groot, 2008; Libben and Titone, 2009). These findings led to the conclusion that language non-selectivity may be restricted by semantically constraining contexts. However, due to the overwhelming body of research supporting language non-selective access (see Brysbaert and Duyck, 2010), a more plausible explanation is that highly predictive target words, cognate and otherwise, will already be activated before a lexical decision needs to be made, thereby diminishing the cognate facilitation effect.

Employing a slightly different paradigm with picture naming instead of lexical decision within sentence context, Starreveld et al. (2014) found that naming latencies were sped up by the sentence context, but cognate effects still remained, especially when naming occurred in L2. In L1, however, the cognate effect was only present in non-predictive sentences. In addition, processing of interlingual homophones also seems unaltered by semantic constraint, even in L1. Lagrou et al. (2013) had Dutch-English bilingual participants perform lexical decision on the last word of a sentence. When this word was an interlingual homophone (e.g., /li:f/: *leaf* - sweet - in Dutch vs. *leaf* in

English), response latencies were slowed down in both language conditions. Although still present, the effect did, however, reduce in size when the homophones were presented in highly semantically constraining sentences. The authors thus concluded that sentence constraint may influence word recognition, but it does not necessarily eliminate cross-lingual lexical interactions.

The story becomes even more interesting when considering studies that have looked at a more natural way of language processing. For instance, Van Assche et al. (2011) recorded Dutch-English bilinguals' eye movements while they read cognates and control words embedded in low and high semantically constraining sentences presented in their second language. Both early and late eye-movement measures yielded cognate facilitation, for low as well as highly constraint sentences. The authors viewed these results as evidence in support of a limited role for top-down influences of semantic constraints on lexical access in both early and later stages of bilingual word recognition. And, more recently, cognate facilitation was even obtained in bilinguals reading an entire book, both for L1 and L2 reading (Cop et al., 2017).

The discrepancies between this and other studies may of course find their origin in the paradigms that are being used. Indeed, when subjects are asked to perform a lexical decision task, response latencies are influenced not only by the time it takes to recognize a word, but also by a decision-making component and even the motor processes required to deliver the manual response (Pinet et al., 2016). Also response strategies that favor either accuracy or speed may play a role. Studies that have compared lexical decision databases with natural reading corpora indeed found that results diverged considerably across paradigms (Kuperman et al., 2013; Dirix et al., 2019). Even across eye tracking corpora correlations of reading times were low whereas within-task reliability was high, illustrating a strong effect of language context. Yet, when aggregating eye tracking measures across multiple representations and contexts, eye tracking measures increasingly converged with lexical decision data, indicating that task-specific language context has a crucial impact on word-level effect manifestation (see Dirix et al., 2019 for a more elaborate view).

## Modeling Bilingual Lexical Access

Theoretical accounts of cross-lingual activation are provided within bilingual language processing models such as the BIA model (Bilingual Interactive Activation model; Dijkstra and Van Heuven, 1998), and its successor BIA+ (Dijkstra and Van Heuven, 2002). These models of word recognition propose that the visual presentation of a word leads to co-activation of many word candidates from different languages that are similar to the input (i.e., orthographic neighbors). This describes the process of language non-selective lexical access. The orthographic representations will subsequently activate their semantic representations (i.e., meaning) (see, for instance, Grainger, 2008) and their phonological representations (e.g., Coltheart et al., 2001). Within this framework, the processing of cognates may be understood by assuming that both orthographic representations of a cognate become activated by the input and subsequently send converging activation to a shared semantic

representation (e.g., Vanlangendonck et al., 2020), which leads to faster processing of the word.

However, in order to provide a general implemented account of word form and meaning retrieval during word production as well as recognition, a localist-connectionist model “Multilink” has recently been developed (Dijkstra et al., 2019). This model integrates the basic assumptions of BIA+ together with the basic architecture of the Revised Hierarchical Model (RHM; Kroll and Stewart, 1994; but see both Brysbaert and Duyck, 2010 and Dijkstra et al., 2019 for a clarification of this model and its issues), and simulates recognition and production of cognates and non-cognates in tasks such as monolingual or bilingual lexical decision, word naming, and word translation production. Multilink is based on a number of assumptions, such as the supposition that the activation of competitors in L1 and L2 directly depends on the orthographic overlap between the input word and stored lexical representations, as operationalized by their Levenshtein distance (Levenshtein, 1966). It also assumes that word candidates compete only at a response choice level but not in terms of lateral inhibition; and that orthographic lexical representations only indirectly (via semantics) cross-linguistically linked. Running simulations on the retrieval of cognates, the model demonstrated that lexical activation spreading from orthographic representations to their (same-language) phonological representations might account for the cognate facilitation effect in word production. This is in line with earlier predictions made by, for instance, Strijkers et al. (2010), who suggested that the overlap between L1 and L2 phonology in cognates results in co-activation of both translation equivalents. This, in turn, culminates in a strong connection between both lexical representations, which is no longer mediated by phonological access. We must, however, note that the current models only simulate single word production and not words within context.

## The Current Study

It is striking that a plethora of research exists on bilingual reading, listening, and speaking, whereas virtually no study has assessed written word production. This mimics the monolingual literature that is also much scarcer in terms of writing, largely because fewer good paradigms exist to assess timing of subprocesses. Still, the same questions pose themselves, such as those on cross-lingual activation. Although spoken and written word production are very distinct activities (i.e., the former requires the utterance of phonemes through vocalization and mouth control, whereas the latter requires the formation of graphemes through control of the hands), lexical access should nevertheless be similar up to a certain stage. Indeed, a basic assumption is that conceptual and lexical processes are shared between spoken and written modalities, whereas post-lexical processes (i.e., phonological/orthographic selection and motor control) are different (e.g., Hillis et al., 1990; Perret and Laganaro, 2013). This implies that also during writing, bilinguals should benefit from a cognate facilitation effect. Taking into account the findings of Starreveld et al. (2014) for spoken production, we might assume that cognate facilitation in L2 written production should even remain present in highly constraining conditions. However, to

the best of our knowledge, this assumption has never been tested. The present study thus set out to examine this issue.

In addition, we wanted to assess the impact of sentence context on such cross-lingual activation. Employing a similar methodology to that of Starreveld et al. (2014), we presented our Dutch-English participants with targets embedded within either low or high constraint sentences formulated in their L2. The targets were cognates and non-cognates presented as pictures, which participants had to name by typing in their responses. If the process of written word production is similar to spoken word production and written/spoken word recognition, a clear cognate facilitation effect should occur, at least in the low constraint context.

## MATERIALS AND METHODS

### Participants

Twenty students in their first Bachelor in Psychology, aged between 18 and 24 ( $M = 18.8$  years; 16 females) participated in the study in exchange for course credit. Prerequisites for participation entailed having Dutch as the native language and a score of at least 70% on the English version of LexTALE (Lemhöfer and Broersma, 2011). Participants were also required to have normal or corrected vision. All individuals signed an informed consent and had the option to exit the experiment at any time of their choosing.

### Materials

One hundred and fifty sentences were presented to participants in their L2 English. These were typewritten, but contained one image presenting one of 25 cognate words, one of 25 matched control words, or one of 25 filler words (see **Supplementary Appendix A**). All images depicting cognates and controls were obtained from the database of Severens et al. (2005). Out of the database containing 590 standardized black-and-white line drawings with picture naming norms in Dutch, we selected those images with a name agreement higher than 75%. Participants were therefore not exposed to the pictures prior to the experiment. Cognate and control words were matched on initial letter, word length (cf. Bates et al., 2003), and word frequency. We employed the SUBTL frequency norms from the SUBTLEXUS corpus (Brysbaert and New, 2009). This procedure provided us with 25 cognates and 25 control words. For a full list of all the words used in the experiment, see **Supplementary Appendix A**. With regard to the cognates, we employed both identical ( $N = 14$ ) and non-identical ( $N = 11$ ) words, with a mean Levenshtein Distance of 0.64 ( $SD = 0.81$ ) based on orthographic overlap.

All sentences had a similar structure, with the target picture presented in the middle of the sentence (see **Table 1** for an example). For a full list of all the sentences used, see **Supplementary Appendix B**. To obscure the purpose of our study, we also included 50 filler sentences. Each picture was presented two times; once in a highly predictive (high constraint) context and once in a non-predictive (low constraint) context. First key stroke latencies were measured (cf. Baus et al., 2013) as

**TABLE 1** | Examples of each condition that will be analyzed.

| Sentence type   | Word type | Example Sentence                                                                                      |
|-----------------|-----------|-------------------------------------------------------------------------------------------------------|
| Low constraint  | Control   | The baron ordered his servant to bring him a PLATE so he could throw it at the wall.                  |
|                 | Cognate   | Jeff is very proud of his PIANO because his grandfather made it for him.                              |
| High constraint | Control   | His mum cooked dinner and put some potatoes and a pork chop on his PLATE before sitting down herself. |
|                 | Cognate   | The white keys are larger than the black keys on a PIANO because they are used more often.            |

the onset of lexical access. There were five blocks of 30 sentences; each containing 10 sentences with a cognate word, 10 sentences with a control word, and 10 sentences with a filler word. All participants saw the 150 sentences in a balanced randomised order and no picture was presented twice in one block.

## Procedure

Participants were seated in front of a computer with their head at approximately 60 cm distance from the monitor. Participants employed a QWERTY keyboard and started the experiment with a short practice phase of 15 warm up trials, containing sentences from all conditions. After the practice phase, the experiment started with instructions presented in the middle of the screen. In addition, verbal instruction was provided. Participants were told to rest their hands in front of the keyboard when they did not have to type. In order to ensure they paid attention to the sentences and not just to the pictures, we clarified that occasionally they would be asked the question: “What was stated in the previous sentence?”. This question appeared on the screen 5 times per block. Participants had to type in their response and press enter to continue the experiment.

After the instructions were read, a fixation cross (+) appeared for 500 ms in the middle of the screen, after which the first word appeared. Participants controlled the speed of reading themselves by pressing the space bar. When the picture was presented on the screen, participants were able to type the name of the depicted object and saw their own text appear directly underneath the picture. They were allowed to correct themselves, as long as they were still on the picture and had not pressed enter. Pressing the enter key led them to the rest of the sentence. After each trial, an empty screen was presented for 1000 ms before the next trial started.

At the end of the experiment, participants were thanked and awarded with course credit. On average, the experiment took about 50–60 min to complete.

## RESULTS

There were 500 observations per condition across participants. No participant answered more than 5 comprehension questions incorrectly, so no participants were excluded from the analysis. The data were trimmed removing incorrect responses (e.g., faulty first strokes) and response times longer than 2.5 standard deviations from the average response time. This procedure eliminated 2.26% of the data, resulting in 350 observations per condition. Incorrect responses in the form of faulty first strokes were common, as the standard use of keyboard in Belgium is not

**TABLE 2** | Means and standard deviations of first stroke latencies (in ms) as a function of sentence constraint and cognate status.

|          | Low Constraint |     | High Constraint |     |
|----------|----------------|-----|-----------------|-----|
|          | Mean           | SD  | Mean            | SD  |
| Controls | 1537           | 234 | 1263            | 144 |
| Cognates | 1327           | 147 | 1363            | 210 |

QWERTY but AZERTY. However, since we removed all matched trials as well (i.e., the word with faulty stroke and its matched cognate/control in both high and low constraint condition), this should not influence our results.

We employed a  $2 \times 2$  factorial design in our experiment. Our variables and their dimensions were Sentence Constraint (Low vs. High) and Word Type (Control vs. Cognate). A two-way repeated measures analysis of variance (ANOVA) was conducted on the means per condition to examine the main effects of Sentence Constraint and the Word Type, and their interaction effect on the first stroke latency of typing a word. We employed a multivariate approach, using Wilks' Lambda. There was a significant main effect of Sentence Constraint [ $F_1(1,19) = 18.31$ ,  $p < 0.001$ ;  $F_2(1,48) = 5.86$ ,  $p = 0.019$ ], with faster responses in high constraint sentences. The main effect of Word Type did not reach significance [ $F_1(1,19) = 3.70$ ,  $p = 0.060$ ;  $F_2(1,48) = 0.10$ ,  $p = 0.758$ ]. There was, however, a significant interaction between Sentence Constraint and Word Type [ $F_1(1,19) = 35.09$ ,  $p < 0.001$ ;  $F_2(1,48) = 9.92$ ,  $p = 0.003$ ], with shorter response latencies to cognates in low constraint sentences and to control words in high constraint sentences.

Further analyses using the results from the multivariate approach revealed that the level of sentence constraint has a significant effect on first stroke latency, but only for control words [ $F_1(1,38) = 22.159$ ,  $p < 0.001$ ;  $F_2(1,23) = 3.80$ ,  $p = 0.001$ ], with shorter latencies in the high constraint condition. In addition, there was a significant effect of Word Type on first stroke latencies, but only in low constraint sentences [ $F_1(1,38) = 12.463$ ,  $p = 0.010$ ;  $F_2(1,24) = 2.55$ ,  $p = 0.018$ ], with shorter response latencies to cognates. Six paired *t*-tests were conducted with the Bonferroni correction for adjusting the significance level to see which of the conditions differed significantly from each other. Only response latencies to control words in the low constraint condition differed from all other conditions. There was no difference between cognate latencies in low versus high constraint sentences. Means, standard deviations and confidence intervals are presented in **Table 2**.

## DISCUSSION

A large body of research suggests that bilingual lexical access is language non-selective (Dijkstra et al., 1999; Costa et al., 2000; Lagrou et al., 2011a). This entails that representations of both languages may be activated simultaneously during word production and recognition, even in a single language context. The cognate facilitation effect has provided evidence for such a hypothesis in written and spoken word recognition as well as in spoken word production. The aim of the current study was to ascertain whether cognate effects also occur in written word production, which has not been previously investigated. It therefore contained pictures of cognate and non-cognate target words embedded within either low or high constraint sentences, which Dutch-English bilinguals had to write in their L2 through typing. First key stroke latencies demonstrated a clear cognate effect for pictures presented in a low constraint condition. That is, cognate words were produced faster than control words, with a mean difference of 210 ms (about 14% faster). However, the cognate effect disappeared in the high constraint sentences, due to the fact that control images were responded to much faster in this condition as compared to the low constraint condition. Response times to cognate images remained the same.

Our results are in line with studies on word recognition (e.g., Schwartz and Kroll, 2006; Van Hell and De Groot, 2008; Libben and Titone, 2009), which reported a cognate facilitation effect for lexical decision in low constraint sentences, but no such effect for high constraint sentences. They are also partially in line with the more similar word production study conducted by Starreveld et al. (2014), as we replicated their findings in a low constraint context, but found no cognate effect in a high constraint context. A possible explanation is that the high constraint sentences in the study by Starreveld et al. were less predictive than those in the current study. Important to note is that Starreveld et al. also found a reduced L2 cognate effect in high constraint context, and no effect whatsoever for L1.

Overall, the results of our study suggest that spoken and written word production are not all that distinct in terms of lexical access. Furthermore, even though response times are longer in general for written production than for spoken production, this may be an artifact of technical specificities, such as visual inspection of the writing. Indeed, a study by Perret and Laganaro (2013) showed that reaction time differences between handwriting and speaking only occur when participants can see and monitor their handwritten production. When they were unable to visually inspect what they were writing, responses took no longer than oral production. Similar conclusion may be drawn for typing, where participants may monitor the position of their fingers on the keys before starting to type.

## REFERENCES

- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychon. Bull. Rev.* 10, 344–380.
- Baus, C., Strijkers, K., and Costa, A. (2013). When does word frequency influence written production? *Front. Psychol.* 4:963. doi: 10.3389/fpsyg.2013.00963
- Blumenfeld, H. K., and Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: examining proficiency and lexical status using eye-tracking. *Lang. Cogn. Proc.* 22, 633–660. doi: 10.1080/01690960601000746
- Brysbaert, M., and Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Biling. Lang. Cogn.* 13, 359–371.
- If we add the current study's outcome to the body of literature on bilingual lexical activation, especially within a constraining sentence context, we may consider two possible explanations for our findings. First of all, it may be the case that sentence context serves as a language cue for appropriate language selection and heightened activation for representations in that language, thereby reducing the cognate effect. Secondly, and more fittingly for the current and previous findings, pre-activation of the target may take place within a highly predictive sentence context before the target stimulus is shown, reducing and even diminishing the cognate effect in high constraint conditions. In other words, pre-activation of the target may be comparable to the activation caused by the presentation itself. This explanation also accounts for the fact that response times on control words are reduced in high constraint conditions, rather than response times on cognates being augmented. Crucially, this explanation fits perfectly within a theory of bilingual access which is language non-selective.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

EW and RC conceptualized the study, with valuable input provided by WD. RC programmed the experiment and collected the data. RC analyzed the results together with EW. EW drafted the manuscript. WD provided critical revisions. All authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.647362/full#supplementary-material>

- Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Brysbaert, M., Van Dyck, G., and Van de Poel, M. (1999). Visual word recognition in bilinguals: evidence from masked phonological priming. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 137–148.
- Caramazza, A., and Brunes, I. (1979). Lexical access in bilinguals. *Bull. Psychon. Soc.* 13, 212–214. doi: 10.3758/BF03335062
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256.
- Cop, U., Dirix, N., Van Assche, E., Drieghe, D., and Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Biling. Lang. Cogn.* 20, 747–769.
- Costa, A., Caramazza, A., and Sebastián Gallés, N. (2000). The cognate facilitation effect: implications for models of lexical access. *J. Exp. Psychol. Learn. Memory Cogn.* 26, 1283–1296. doi: 10.1037/0278-7393.26.5.1283
- Dijkstra, A., and Van Heuven, W. J. B. (1998). “The BIA-model and bilingual word recognition,” in *Localist Connectionist Approaches to Human Cognition*, eds J. Grainger and A. Jacobs (Hillsdale, NJ: Lawrence Erlbaum Associates), 189–225.
- Dijkstra, T., Timmermans, M., and Schriefers, H. (2000). On being blinded by your other language: effects of task demands on interlingual homograph recognition. *J. Memory Lang.* 42, 445–464.
- Dijkstra, T., and Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingualism* 5, 176–197. doi: 10.1017/S1366728902003012
- Dijkstra, T., Van Jaarsveld, H., and Ten Brinke, S. (1998). Interlingual homograph recognition: effects of task demands and language intermixing. *Biling. Lang. Cogn.* 1, 51–66.
- Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., et al. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Biling. Lang. Cogn.* 22, 657–679.
- Dijkstra, T., Grainger, J., and Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *J. Memory Lang.* 41, 496–518. doi: 10.1006/jmla.1999.2654
- Dirix, N., Brysbaert, M., and Duyck, W. (2019). How well do word recognition measures correlate? Effects of language context and repeated presentations. *Behav. Res. Methods* 51, 2800–2816.
- Duyck, W., Van Assche, E., Drieghe, D., and Hartsuiker, R. (2007). Visual word recognition by bilinguals in a sentence context: evidence for non-selective lexical access. *J. Exp. Psychol. Learn. Memory Cogn.* 33, 663–679.
- Grainger, J. (2008). Cracking the orthographic code: an introduction. *Lang. Cogn. Processes* 23, 1–35.
- Hillis, A. E., Rapp, B. C., Romani, C., and Caramazza, A. (1990). Selective impairment of semantics in lexical processing. *Cogn. Neuropsychol.* 7, 191–224.
- Hoshino, N., and Kroll, J. F. (2008). Cognate effects in picture naming: does cross-language activation survive a change of script? *Cognition* 106, 501–511. doi: 10.1016/j.cognition.2007.02.001
- Jared, D., and Szucs, C. (2002). Phonological activation in bilinguals: evidence from interlingual homograph naming. *Biling. Lang. Cogn.* 5, 225–239.
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Memory Lang.* 33, 149–174.
- Kuperman, V., Drieghe, D., Keuleers, E., and Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Q. J. Exp. Psychol.* 66, 563–580.
- Lagrou, E., Hartsuiker, R. J., and Duyck, W. (2011a). Knowledge of a second language influences auditory word recognition in the native language. *J. Exp. Psychol. Learn. Memory Cogn.* 37, 952–965. doi: 10.1037/a0023217
- Lagrou, E., Hartsuiker, R. J., and Duyck, W. (2011b). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Biling. Lang. Cogn.* 16, 508–517. doi: 10.1017/S1366728912000508
- Lagrou, E., Hartsuiker, R. J., and Duyck, W. (2013). Interlingual lexical competition in a spoken sentence context: evidence from the visual world paradigm. *Psychon. Bull. Rev.* 20, 963–972.
- Lagrou, E., Hartsuiker, R. J., and Duyck, W. (2015). Do semantic constraint and L2 proficiency influence language selectivity of lexical access in native language listening? *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1524–1538.
- Lemhöfer, K., and Broersma, M. (2011). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behav. Res. Methods* 44, 325–343.
- Lemhöfer, K., and Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: effects of code similarity in language-specific and generalized lexical decision. *Memory Cogn.* 2, 533–550.
- Lemhöfer, K., Dijkstra, T., and Michel, M. C. (2004). Three languages, one ECHO: cognate effects in trilingual word recognition. *Lang. Cogn. Processes* 19, 585–611. doi: 10.1080/01690960444000007
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions insertions and reversals. *Cybernet. Control Theory* 10, 707–710.
- Libben, M. R., and Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *J. Exp. Psychol. Learn. Memory Cogn.* 35, 381–390. doi: 10.1037/a0014875
- Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition* 25, 71–102.
- Perret, C., and Laganaro, M. (2013). Why are written picture naming latencies (not) longer than spoken naming? *Read. Writing* 26, 225–239.
- Pinet, S., Dubarry, A. S., and Alario, F. X. (2016). Response retrieval and motor planning during typing. *Brain Lang.* 159, 74–83.
- Schwartz, A. I., and Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *J. Memory Lang.* 55, 197–212. doi: 10.1016/j.jml.2006.03.004
- Schwartz, A. I., Kroll, J. F., and Diaz, M. (2007). Reading words in Spanish and English: mapping orthography to phonology in two languages. *Lang. Cogn. Processes* 22, 106–129. doi: 10.1080/01690960500463920
- Severens, E., Van Lommel, S., Ratinckx, E., and Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychol.* 119, 159–187. doi: 10.1016/j.actpsy.2005.01.002
- Spivey, M., and Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychol. Sci.* 10, 281–284. doi: 10.1017/S1366728903001068
- Starreveld, P. H., de Groot, A. M. B., Rossmark, B. M. M., and van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: evidence from word production in sentence context. *Biling. Lang. Cogn.* 17, 258–276. doi: 10.1111/1467-9280.001519280.00151
- Strijkers, K., Costa, A., and Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cereb. Cortex* 20, 912–928.
- Van Assche, E., Drieghe, D., Duyck, W., Welsaert, M., and Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *J. Memory Lang.* 64, 88–107.
- Van Assche, E., Duyck, W., Hartsuiker, R., and Diependaele, K. (2009). Does bilingualism change native-language reading? *Cognate effects in sentence context. Psychol. Sci.* 20, 923–927.
- Van Hell, J. G., and De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychol.* 128, 431–451. doi: 10.1016/j.actpsy.2008.03.010
- Vanlangendonck, F., Peeters, D., Rueschemeyer, S.-A., and Dijkstra, T. (2020). Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Biling. Lang. Cogn.* 23, 836–844.
- Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Memory Lang.* 50, 1–25.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Woumans, Clauws and Duyck. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# What Are the Modulators of Cross-Language Syntactic Activation During Natural Reading?

Naomi Vingron<sup>1\*</sup>, Pauline Palma<sup>1</sup>, Jason W. Gullifer<sup>1</sup>, Veronica Whitford<sup>2</sup>, Deanna Friesen<sup>3</sup>, Debra Jared<sup>4</sup> and Debra Titone<sup>1\*</sup>

<sup>1</sup>Department of Psychology, McGill University, Montreal, QC, Canada, <sup>2</sup>Department of Psychology, University of New Brunswick, Fredericton, NB, Canada, <sup>3</sup>Faculty of Education, University of Western Ontario, London, ON, Canada, <sup>4</sup>Department of Psychology, University of Western Ontario, London, ON, Canada

## OPEN ACCESS

### Edited by:

Zofia Wodniecka,  
Jagiellonian University, Poland

### Reviewed by:

Evy Adèle Woumans,  
Ghent University, Belgium  
Aleksandra Tomic,  
Arctic University of Norway, Norway

### \*Correspondence:

Naomi Vingron  
naomi.vingron@mail.mcgill.ca  
Debra Titone  
debra.titone@mcgill.ca

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 21 August 2020

**Accepted:** 27 May 2021

**Published:** 06 July 2021

### Citation:

Vingron N, Palma P, Gullifer JW,  
Whitford V, Friesen D, Jared D and  
Titone D (2021) What Are the  
Modulators of Cross-Language  
Syntactic Activation During Natural  
Reading?  
Front. Commun. 6:597701.  
doi: 10.3389/fcomm.2021.597701

Bilinguals juggle knowledge of multiple languages, including syntactic constructions that can mismatch (e.g., the red car, la voiture rouge; Mary sees it, Mary le voit). We used eye-tracking to examine whether French-English ( $n = 23$ ) and English-French ( $n = 21$ ) bilingual adults activate non-target language syntax during English L2 (Experiment 1) and L1 (Experiment 2) reading, and whether this differed from functionally monolingual English reading (Experiment 3,  $n = 26$ ). People read English sentences containing syntactic constructions that were either partially shared across languages (adjective-noun constructions) or completely unshared (object-pronoun constructions). These constructions were presented in an intact form, or in a violated form that was French-consistent or French-inconsistent. For both L2 and L1 reading, bilinguals read French-consistent adjective-noun violations relatively quickly, suggesting cross-language activation. This did not occur when the same people read object-pronoun constructions manipulated in the same manner. Surprisingly, English readers exposed to French in their lifetime but functionally monolingual, also read French-consistent violations for adjective-noun constructions faster, particularly for some items. However, when we controlled for item differences in the L2 and L1 reading data, cross-language effects observed were similar to the original data pattern. Moreover, individual differences in L2 experience modulated both L2 and L1 reading for adjective-noun constructions, consistent with a cross-language activation interpretation of the data. These findings are consistent with the idea of syntactic cross-language activation during reading for some constructions. However, for several reasons, cross-language syntactic activation during comprehension may be overall more variable and challenging to investigate methodologically compared to past work on other forms of cross-language activation (i.e., single words).

**Keywords:** bilingualism, syntax, cross-language activation, eye movements, reading

## INTRODUCTION

What are the modulators of cross-language syntactic activation during natural reading? Bilinguals juggle multiple languages in everyday communication, yet their ability to produce and comprehend usually proceeds fluently (Grosjean, 2001). Consider the following sentence, “My neighbors had a heated chat about the ousted man’s strange Tweets.” While interpretation of this sentence is

straightforward, there are several places where a French-English bilingual may have difficulty. First, the word “chat,” an interlingual homograph, could simultaneously activate the English meaning “informal conversation” or the French meaning “cat.” We know from many studies (reviewed in Van Assche et al., 2012; Lauro and Schwartz, 2017; Palma and Titone, 2020), and leading models of bilingual language processing (e.g., Dijkstra and Van Heuven, 2002; Dijkstra, et al., 2019), that cross-language activation of divergent meanings slows reading for words like “chat.” While much is understood about cross-language activation at the single-word level during reading (reviewed in Jared, 2015; Titone et al., 2016; Palma and Titone, 2020), less clear is whether bilingual adults experience cross-language activation for multiword syntactic constructions during reading (see Roberts, 2012, for a review) and whether this activation is modulated by individual differences among bilinguals.

Prior eye-tracking studies of sentence reading (e.g., Titone et al., 2011; Pivneva et al., 2014; Whitford and Titone, 2015; Friesen et al., 2020) and paragraph reading (e.g., Whitford and Titone, 2012; Whitford and Titone, 2015) have highlighted the influence of current L2 exposure on lexical access and reading fluency. Many studies reported bidirectional L1-L2 influences that are modulated by individual differences in current L2 exposure, although these studies were not focused on syntactic processing per se. It is therefore unclear whether, and to what extent, individual differences in bilingual experience modulate cross-language syntactic activation. On the one hand, some studies found that cross-language syntactic activation decreases as L2 experience increases, as individuals rely less on knowledge from their L1 (e.g., Dussias and Sagarra, 2007; Dussias et al., 2015; Kasprian and Steinhauer, 2017; see also; Roberts, 2012). On the other hand, some studies found that greater L2 proficiency is associated with larger cross-language activation effects, as a consequence of more integrated syntactic processing (e.g., Berne et al., 2009). Complicating matters is that syntactic processing is variable, even in monolingual individuals. Although formal linguistic approaches predict minimal individual differences in native language syntactic processing (see Kidd et al., 2018), individual differences have been observed both behaviorally and neurally (e.g., Wells et al., 2009; Street and Dabrowska, 2014; Mahowald and Fedorenko, 2016).

To the extent that cross-language activation occurs for syntactic processing, the comprehension of English adjective-noun constructions, such as “interesting chat” and “strange Tweets,” might be affected by knowledge of French because the reverse word order is more typical of French: “une conversation intéressante” or “Tweets étranges.” Similarly, the comprehension of object-pronoun constructions, such as “Mary sees it” and “They love her,” might also be influenced by knowledge of French because object-pronouns are systematically cliticized to the verb in French: “Mary le voit” or “Ils l’aiment.” In contrast, English object-pronouns are systematically placed after the verb (e.g., “Mary sees it”). It should be noted that the word order of French adjective-noun constructions is aligned with English for some adjectives (e.g., “la

nouvelle maison”), whereas word order in object-pronoun constructions is always different in the two languages.

More generally, adjectives are linguistically optional, in that their omission neither impedes comprehension nor affects grammaticality (e.g., “the balloon”/“le ballon”), whereas omission of object-pronouns in transitive constructions causes ungrammaticality (e.g., “\*Mohammed washes”/“\*Mohammed lave”). As such, object-pronouns constructions also differ from adjective-noun constructions in their morphosyntactic behavior across languages. Specifically, English object-pronouns are strong pronouns, thus functioning syntactically as a lexical determiner phrases (DPs; Cardinaletti and Starke, 1999), whereas French object-pronouns are clitics, a class of pronouns that differs semantically, morphologically and syntactically from strong pronouns (Cardinaletti and Starke, 1999). In contrast to strong pronouns, clitics never occur in isolation (e.g., “Qui as-tu vu Mary embrasser hier? \*La”), cannot be coordinated (e.g., “J’ai vu \*la et Mary s’embrasser”) or be modified by adverbs (e.g., “Seule \*la est assez rapide”). Such differences between the two constructions can potentially impact the relative degree of cross-language activation for the same bilingual readers.

Here, we investigated whether bilingual adults, whose first language was French or English, would experience cross-language syntactic activation during sentence reading for these two types of constructions. To investigate this issue, we used a procedure in which English sentences containing these constructions violated English grammar in a manner that was either consistent or inconsistent with French. Specifically, we created adjective-noun constructions that were consistent with French adjective-noun word order (e.g., “The man saw the vehicle German that was parked on the street.”), and sentences containing adjective-noun violations that were inconsistent with French adjective-noun word order (e.g., “The man saw German the vehicle that was parked on the street.”). We also created English object-pronoun constructions that were consistent with French object-pronoun word order (e.g., “Leah baked the birthday cake, and she it ate with all her friends”), and sentences containing object-pronoun constructions that were inconsistent with French object-pronoun word order (e.g., “Leah baked the birthday cake, and ate she it with all her friends”).

To generate predictions about how people would respond to these grammatical violations, we turned to a prominent model of bilingual sentence processing, the Unified Competition model (UCM; MacWhinney, 2005). It posits that a bilingual’s two languages are co-activated and compete for selection to the degree that they mismatch cross-linguistically (MacWhinney, 1987; Freck-Mestre, 2005; Kroll and Tokowicz, 2005). The UCM predicts that co-activation of similar L1/L2 syntactic constructions should lead to minimal competition and possibly facilitation during L2 reading—a phenomenon called positive transfer. In contrast, co-activation of mismatching syntactic constructions should block positive transfer from L1 to L2, causing L1 activation to impede L2 production or comprehension, and, ultimately, result in cross-language competition (e.g., Tuninetti et al., 2015).

Applied to the experimental manipulation here, the UCM would predict greater tolerance for L2 syntactic violations that are

consistent with L1 French syntax; violations of English sentences that are consistent with French should be read more easily by French L1 participants than violations that are not consistent with French. According to the UCM, both adjective-noun and pronoun constructions differ across English and French, which should lead to greater tolerance of French-consistent violations. It is important to underline that the UCM formulates predictions based on offline behaviors (i.e., grammaticality judgments) and does not make explicit predictions about online behaviors (i.e., eye movements during reading). Finally, the UCM makes predictions about bilinguals in general and not about highly proficient bilinguals in particular, such as the groups tested in this study. In addition, the model does not make predictions about how L2 knowledge may influence L1 syntactic processing. Nevertheless, as mentioned above, several studies involving highly proficient bilinguals have suggested that there is a bidirectional effect of a bilingual's two languages on processing (Van Assche et al., 2009; Titone et al., 2011). Thus, these models remain useful for the purpose of this work.

Very few studies have investigated syntactic-level cross-language activation, and the ones that exist are somewhat variable in outcome. In an event-related potential (ERP) go-no go study, Thierry and Sanoudaki (2012) found that early Welsh-English bilinguals exhibited a modulation of the amplitude of the N2 component when reading English adjective-noun constructions consistent with Welsh word order (e.g., \*the book red).<sup>1</sup> This effect, which was not found in English monolingual participants, was interpreted as a consequence of the activation of L1 Welsh grammar during L2 English processing. In an eye-tracking grammaticality judgment study that inspired this investigation, Tuninetti et al. (2015) compared English monolinguals with different groups of bilinguals whose L1 was either consistent with English in terms of adjective-noun word order (Mandarin-English bilinguals) or inconsistent (Arabic-English bilinguals). Critical ungrammatical English sentences varied in their compatibility with these different L1s. On early reading measures, all three groups were equally tolerant of adjective-noun word order violations, suggesting no cross-language syntactic activation. However, they differed in how long it took to repair the violations (i.e., later reading measures).

<sup>1</sup>For participants, a go-no go task consists in responding to some visual stimuli (go condition), but to refrain from responding to others (no go condition). A modulation of the N2 component is classically observed in the *no go*, but not in the *go* condition, suggesting that this component is associated with cognitive inhibition (e.g., Falkenstein et al., 1999). Thierry and Sanoudaki (2012) instructed participants to respond only to adjective-noun constructions that corresponded—at least partially—to presented pictures (e.g., blue book, picture of a red book), and to refrain from responding to constructions inconsistent with presented pictures (e.g., green car, picture of a red book). The authors hypothesized that monolingual participants would exhibit a modulation of the N2 when expecting a noun after an adjective (e.g., blue book), whereas the same participants would not expect an adjective after a noun (e.g., \*car red). In contrast, bilingual participants were hypothesized to expect both a noun after an adjective (consistent with English grammar) and an adjective following a noun (consistent with Welsh grammar), resulting in a modulation of the N2 component in both cases.

Although important, the conclusions from these studies are somewhat limited, as they only investigated one type of syntactic construction (adjective-noun constructions), limiting their generality. The use of explicit paradigms (binary decision paradigms, grammaticality judgments) may also have limited the naturalness of their task. Moreover, inspection of Tuninetti et al. (2015) materials reveals that experimental sentences containing violations may not have been matched in other ways, such as the degree of word transpositions across the conditions (i.e., L1 consistent or inconsistent), which is known to impact whether readers even notice word order violations in monolingual reading (reviewed in Snell and Grainger, 2019). Furthermore, bilingual readers may be more likely to tolerate the cross-language condition, not because of cross-language syntactic activation per se, but rather because they are more susceptible to making word transposition errors (as a result of limited proficiency) during comprehension.

For these reasons, increasing the validity of a cross-language syntactic activation experiment would require all cross-language violations (and control violations) to be matched on word transposition characteristics (as well as cues to ungrammaticality, which are likely correlated). It may also be important to examine how the same people read multiple constructions that systematically differ across languages in a manner that could lead to more or less cross-language activation, such as adjective-noun and object-pronoun constructions, even if those two constructions cannot be statistically compared directly because of the myriad ways they differ in a low-level sense (e.g., length and type of words, overall frequency, the kinds of sentence frames in which they are embedded, likelihood of differential parafoveal preview, etc.). Finally, to the extent that cross-language activation results in higher tolerance of French-consistent violations (i.e., faster reading times compared to a French-inconsistent violation), we would expect that reading performance would be modulated by individual differences in bilingual experience, specifically the amount of L2 usage and the likelihood to find oneself in a setting where the two languages are mixed (i.e., in Montreal, both English and French are often used interchangeably in downtown restaurants and stores).

## The Present Study

With the above logic in mind, we examined how bilingual adults read sentences that contained word order violations to assess cross-language syntactic activation. Like Tuninetti et al. (2015), we used eye-tracking. Unlike Tuninetti et al. (2015), participants simply read sentences for comprehension rather than making explicit grammaticality judgments. Additionally, object-pronoun constructions were studied for the same people, using the same procedure, alongside adjective-noun constructions.

We posed three main questions: 1) Do bilingual readers show evidence of cross-language syntactic activation during L2 and L1 reading of adjective-noun and object-pronoun constructions? 2) Do individual differences in language mixing and general frequency of L2 speaking modulate cross-language activation patterns during bilingual reading? and 3) How do monolingual reading patterns of adjective-noun and object-pronoun constructions compare to those of bilinguals?

To answer these questions, we conducted three experiments. Experiment 1 tested French-English bilingual adults who read English sentences containing the grammatical violations described above. Experiment 2 tested English-French bilingual adults who read the same English sentences. Experiment 3 tested functionally monolingual English-speaking adults on the same sentences. Across Experiments 1 and 2, we also examined how graded differences in bilingual experience (e.g., general frequency of L2 speaking and degree of language mixing) modulated cross-language activation patterns. Based on previous studies, we further predicted both L2 and L1 reading of these constructions would be modulated by individual differences in language exposure.

## EXPERIMENT 1: FRENCH-ENGLISH BILINGUALS READING IN ENGLISH (L2 READING)

### Method

**Participants.** We tested 25 French-English bilinguals (6 men, 19 women) at McGill University and the University of Western Ontario, who did not speak additional languages fluently according to self-report. We tested across two different sites with the hope of maximizing our chances of having a large range of French-English bilinguals with respect to use of English (their L2).

Participants had a mean age of 21.36 years ( $SD = 2.37$ ), with no uncorrected vision, speech, or reading impairments. They were recruited through online advertisements, local posters, and the Psychology Department subject pool. Subject pools participants were given course credit, while the other participants were paid \$10/hour.

All participants completed a language history questionnaire (Marian et al., 2007), which included self-reported frequency of reading, writing, listening, and speaking in L1 and L2 (where applicable), language mixing, and acquisition history. General frequency of L1 and L2 speaking, listening, writing, and reading were reported in percentages (e.g., What percentage of the time do you speak French?). Participants reported speaking their L2, English, 57.83% ( $SD = 23.60$ ; Min = 20; Max = 90) of the time. On average, they rated their likelihood to mix French and English (e.g., situations where the two languages can be used in a complementary or interchangeable way) as 4.60 out of 7 ( $SD = 1.41$ ; Min = 2; Max = 7).<sup>2</sup>

**Materials.** Materials consisted of 63 sentences containing adjective-noun constructions and 54 sentences containing object-pronoun constructions. Of note, three of the items containing adjective-noun constructions were excluded from the analyses because the manipulation did not appear as intended (e.g., “The

picture captured the [father’s proud smile] that cheered the runner.” when presented in the French-inconsistent condition read “The picture captured the [proud father’s smile] that cheered the runner”, which is grammatical). Assignment of sentences to experimental conditions was counterbalanced so that each participant only saw each sentence in one condition. Exemplar sentences for each experimental condition are presented in **Table 1**. In addition to the experimental sentences, materials also included 78 sentences containing a verb particle construction that appeared either in intact or in violated form. As well, we included 90 random filler sentences, of which ten were ungrammatical jabberwocky sentences.

In creating the experimental sentences containing adjective-noun and object-pronoun constructions, our aim was to minimize bias and ensure that the manipulation was as natural sounding as possible. Importantly, the creation of materials for this experiment was guided by well-established findings regarding eye movements during sentence reading (see Clifton et al., 2007). As such, the experimental sentences were consistent in length and complexity. The region of interest containing the manipulated syntactic construction was always placed in the middle of the sentence and was never immediately followed by any punctuation so as to avoid sentence wrap-up effects (Hirotani et al., 2006).

**Procedure.** Participants first completed the eye-tracking reading task, followed by the language background questionnaire. For the reading task, participants read each sentence (presented one at a time) silently for comprehension and indicated *via* button-press when they finished reading each one. Sentences were displayed in 10-point, yellow Monaco font on a black background. Participants were asked to respond to 21 yes/no comprehension questions on filler trials to ensure participants were attentive throughout the experiment.

**Apparatus.** Eye-movement data were acquired at a rate of 1,000 Hz from the right eye using an EyeLink 1,000 desktop mounted system (SR-Research, Ontario, Canada). Sentences were displayed on a 21-inch ViewSonic CRT monitor, positioned 57 cm from the participant. We presented sentences using UMass EyeTrack software (downloadable from: <https://blogs.umass.edu/eyelab/software/>). Participants eyes were calibrated using a 9-point calibration. On each trial, a gaze contingent yellow box was presented on the left of the screen before the sentence appeared.

**Data preprocessing and analytic approach.** We used the UMass Amherst EyeTrack software to manually clean and extract eye movement data. We first viewed each trial in the EyeDoctor program and removed blinks and trials containing evidence of track loss. Finally, we used EyeDry to extract the eye movement reports that would be used for analysis.

We analyzed both early and late eye movement measures for the entire determiner-adjective-noun region and object-pronoun region. Specifically, we considered first pass gaze duration, which describes the amount of time (in ms) that the eye is in a critical region before exiting it to the right for the first time and total reading time, which refers to the total amount of time the eye spends in a critical region during a trial, including regressions. For example, in the sentence “The man saw the German vehicle that was parked on the street,” containing an adjective-noun construction, the region of interest was defined as “the German vehicle” In the sentence “Leah baked the birthday cake, and she

<sup>2</sup>Of note, the language mixing variable used here differs from a more traditional measure of code switching. Specifically, code switching usually refers to bilingual speakers inserting single words or phrases from one language into utterances in another language or alternating between different languages from one sentence to the next (Green and Wei, 2014; Green and Wei, 2016). Language mixing is understood as a more general measure that refers to situations in which a bilingual individual may use their languages in a more integrated or complementary way.

**TABLE 1 |** Sample sentences containing adjective-noun and object-pronoun constructions across conditions.

| Sentence type       | Adjective-noun constructions                                  | Object-pronoun constructions                                                           |
|---------------------|---------------------------------------------------------------|----------------------------------------------------------------------------------------|
| English intact      | The man saw the German vehicle that was parked on the street. | Mark orders a chicken, and <i>he serves it</i> to his friends while they watch a movie |
| French-consistent   | The man saw the vehicle German that was parked on the street. | Mark orders a chicken, and <i>he it serves</i> to his friends while they watch a movie |
| French-inconsistent | The man saw German the vehicle that was parked on the street. | Mark orders a chicken, and serves <i>he it</i> to his friends while they watch a movie |

ate it with all her friends,” which contains an object-pronoun construction, the region of interest was defined as “she ate it.”

Included in the analysis were only observations with first pass gaze durations or reading times on the three-word region of interest lasting longer than 80 ms in total. When fixation durations were below this cut-off, the region was considered skipped. In the adjective-noun experiment, this led to the exclusion of six observations from the first pass gaze duration analysis and three observations from the total reading time analysis.<sup>3</sup> In the object-pronoun experiment, this led us to exclude sixteen observations from the first pass gaze duration analysis and eight observations from the total reading time analysis.<sup>4</sup> An upper cut-off of 10,000 ms was applied, but no observations exceeded this.

First pass gaze duration and total reading time data were log-transformed, analyzed using linear mixed-effects (LME) models, and plotted in R (R Core Team, 2017) using the following packages: lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), ggplot2 (Wickham, 2016), and effects (Fox and Weisberg, 2018). We first computed a set of core models, where the goal was to assess the core manipulation of the experiment (i.e., French-consistent and -inconsistent manipulations) across all participants. Here, the categorical independent variable (IV), sentence type, was treatment coded (0, 1), where the French-consistent sentence type served as the baseline against comparisons with the English intact and French-inconsistent sentence types. In all core models, trial order (continuous, z-scored) was a control variable, and we used maximal random effects as supported by the data. When a model did not converge, we followed the procedures outlined in Barr et al. (2013) to simplify the random effects structure.

We next fit separate models to evaluate whether individual differences among bilinguals in either the general frequency of L2 speaking (which we took as a general usage measure) or language mixing interacted with sentence type across the different reading measures (i.e., gaze duration and total reading time of the region). We computed an additional model for each measure that included the interaction of language mixing (continuous, z-scored) with sentence type, as well as trial order and general frequency of L2 speaking as

control variables and using random intercepts only. We used the same procedure to investigate the impact of general frequency of L2 speaking in an interaction with sentence type, in which we also controlled for language mixing and trial order.

Across all models, we evaluated significance using Satterthwaite approximations, implemented in the lmerTest package (Kuznetsova et al., 2017). **Table 2** and **Table 3** present the descriptive statistics for the eye movement measures. Full datasets and output of all subsequently reported models are available from the OSF repository ([https://osf.io/jec5s/?view\\_only=a9d0ad4f9b994cd9b93371d2a2089cf1](https://osf.io/jec5s/?view_only=a9d0ad4f9b994cd9b93371d2a2089cf1)).

## Results

**Sentence comprehension.** Sentence comprehension accuracy was 90% among French-English bilinguals, indicating that they were attentive during the reading task.

### Adjective-Noun Constructions

**Core models.** A treatment coded model with French-consistent as the baseline showed that constructions embedded in French-consistent sentences and English intact sentences had similar first pass gaze durations ( $\beta = -0.01$ ,  $SE = 0.02$ ,  $t = -0.70$ ,  $p = 0.48$ ). In contrast, constructions embedded in French-consistent sentences had significantly shorter first pass gaze durations compared than those embedded in French-inconsistent sentences ( $\beta = 0.07$ ,  $SE = 0.02$ ,  $t = 3.41$ ,  $p < 0.01$ ).<sup>5</sup>

For total reading time, a treatment coded model with French-consistent as the baseline showed that constructions embedded in French-consistent sentences were read significantly faster than those embedded in French-inconsistent sentences ( $\beta = 0.15$ ,  $SE = 0.02$ ,  $t = 5.69$ ,  $p < 0.01$ ), but slower than English intact sentences ( $\beta = -0.13$ ,  $SE = 0.02$ ,  $t = -4.94$ ,  $p < 0.01$ ).<sup>6</sup>

**Effects of bilingual language experience.** We next evaluated whether performance interacted with individual differences in L2 (English) experience (i.e., language mixing and general frequency of L2 speaking). Accordingly, we reran the above models including general frequency of L2 speaking as an interaction term and language mixing as a control variable, and vice versa. We discuss only significant interaction effects with these variables if they also involve condition. Across these models, individual

<sup>3</sup>Of the observations excluded from the first pass gaze duration analysis, three were presented in the English intact condition, two in the French-consistent and one in the French-inconsistent condition. Of those, only the observations presented in the French-consistent and French-inconsistent conditions were also excluded from the total reading time analysis.

<sup>4</sup>Of the observations excluded from the first pass gaze duration analysis, five were presented in the English intact condition, four in the French-consistent and seven in the French-inconsistent condition. Of those, two of the observations presented in the English intact condition, three of the French-consistent and three of the observations presented in the French-inconsistent condition were also excluded from the total reading time analysis.

<sup>5</sup>A refit model with the reference level set at French-inconsistent showed that first pass gaze durations for French-inconsistent constructions were longer than those embedded in English intact constructions ( $\beta = -0.09$ ,  $SE = 0.02$ ,  $t = -4.12$ ,  $p < 0.01$ ).

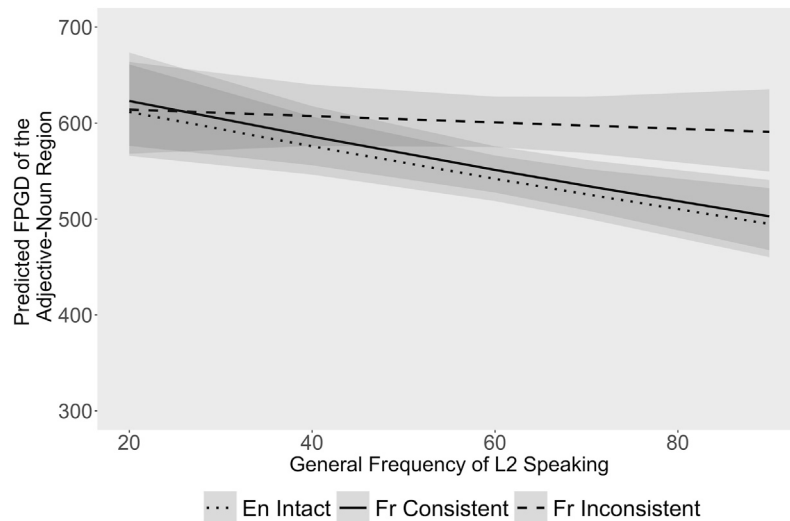
<sup>6</sup>A refit model with the reference level set at French-inconsistent showed that total reading time for French-inconsistent constructions were longer than those embedded in English intact constructions ( $\beta = -0.29$ ,  $SE = 0.02$ ,  $t = -10.65$ ,  $p < 0.01$ ).

**TABLE 2 |** Descriptive statistics for all dependent measures of adjective-noun constructions for Experiment 1 (left), Experiment 2 (center) and Experiment 3 (right).

| Sentence type                        | Adjective-noun constructions           |     |       |     |                                        |     |      |     |                      |     |      |     |
|--------------------------------------|----------------------------------------|-----|-------|-----|----------------------------------------|-----|------|-----|----------------------|-----|------|-----|
|                                      | French-English bilinguals (L2 reading) |     |       |     | English-French bilinguals (L1 reading) |     |      |     | English monolinguals |     |      |     |
|                                      | FPGD                                   |     | TRT   |     | FPGD                                   |     | TRT  |     | FPGD                 |     | TRT  |     |
|                                      | Mean <sup>a</sup>                      | SD  | Mean  | SD  | Mean                                   | SD  | Mean | SD  | Mean                 | SD  | Mean | SD  |
| English intact (the sunny room)      | 591                                    | 228 | 709   | 336 | 511                                    | 244 | 614  | 358 | 523                  | 216 | 605  | 396 |
| French-consistent (the room sunny)   | 605                                    | 246 | 850   | 512 | 551                                    | 255 | 831  | 538 | 552                  | 241 | 746  | 424 |
| French-inconsistent (sunny the room) | 647                                    | 251 | 1,000 | 611 | 583                                    | 298 | 962  | 665 | 593                  | 279 | 867  | 527 |

<sup>a</sup>All means and SD are in ms.**TABLE 3 |** Descriptive statistics for all dependent measures of object-pronoun constructions for Experiment 1 (left), Experiment 2 (center) and Experiment 3 (right).

| Sentence type                    | Object-pronoun constructions           |     |      |     |                                        |     |      |     |                      |     |      |     |
|----------------------------------|----------------------------------------|-----|------|-----|----------------------------------------|-----|------|-----|----------------------|-----|------|-----|
|                                  | French-English bilinguals (L2 reading) |     |      |     | English-French bilinguals (L1 reading) |     |      |     | English monolinguals |     |      |     |
|                                  | FPGD                                   |     | TRT  |     | FPGD                                   |     | TRT  |     | FPGD                 |     | TRT  |     |
|                                  | Mean                                   | SD  | Mean | SD  | Mean                                   | SD  | Mean | SD  | Mean                 | SD  | Mean | SD  |
| English intact (she ate it)      | 421 <sup>a</sup>                       | 224 | 499  | 262 | 417                                    | 204 | 476  | 274 | 387                  | 183 | 442  | 226 |
| French-consistent (she it ate)   | 521                                    | 304 | 759  | 535 | 464                                    | 340 | 659  | 559 | 427                  | 211 | 575  | 388 |
| French-inconsistent (ate she it) | 508                                    | 298 | 749  | 628 | 446                                    | 261 | 720  | 556 | 448                  | 243 | 625  | 441 |

<sup>a</sup>All means and SD are in ms.**FIGURE 1 |** Experiment 1 (French-English bilinguals reading in L2), interaction effect of general frequency of L2 speaking (out of 100%) on predicted first pass gaze duration (in ms) of the adjective-noun region. Shaded area represents plus/minus one standard error of the mean.

differences in general frequency of L2 speaking interacted with the French-inconsistent condition for first pass gaze duration. Specifically, there was a significant interaction between the French-inconsistent vs. French-consistent contrast and general frequency of L2 speaking ( $\beta = 0.06$ ,  $SE = 0.02$ ,  $t = 2.70$ ,  $p < 0.01$ ;

see **Figure 1**). When the baseline was set to French-inconsistent, there was a significant interaction between the French-inconsistent vs. English intact contrast and general frequency of L2 speaking ( $\beta = -0.05$ ,  $SE = 0.02$ ,  $t = -2.67$ ,  $p < 0.01$ ). This suggests that when L2 experience is low, French-English bilingual participants do not

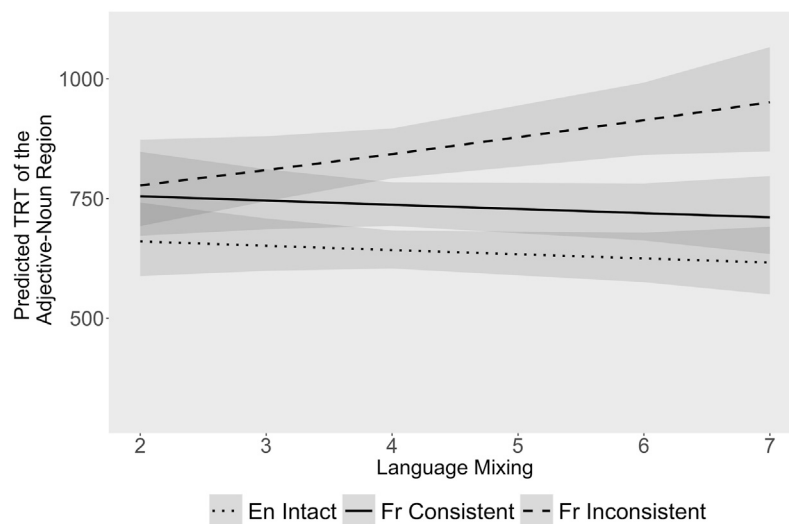


**FIGURE 2 |** Experiment 1 (French-English bilinguals reading in L2), interaction effect of language mixing (out of 7) on predicted total reading time (in ms) of the adjective-noun region. Shaded area represents plus/minus one standard error of the mean.

distinguish between the three sentence types. As L2 experience increases, readers are able to differentiate between French-inconsistent ungrammatical sentences and the other two sentence types (which were processed more readily). Importantly, there were no differences between French-consistent sentences and English intact sentences.

In contrast with L2 speaking that impacted first pass gaze duration, individual differences in language mixing impacted total reading time. When the baseline was set to French-consistent, we found a significant interaction between the French-inconsistent vs. French-consistent contrast and language mixing ( $\beta = 0.07$ ,  $SE = 0.02$ ,  $t = 2.69$ ,  $p < 0.01$ ; see **Figure 2**). When the baseline was set to French-inconsistent, there was a significant interaction between the French-inconsistent vs. English intact contrast and language mixing ( $\beta = 0.07$ ,  $SE = 0.02$ ,  $t = -2.79$ ,  $p < 0.01$ ). These significant interactions suggest that a higher frequency of mixing French and English was associated with a greater difference in total reading times between English intact constructions and French-inconsistent constructions, as well as between French-consistent constructions and French-inconsistent constructions.

### Object-Pronoun Constructions

**Core models.** A treatment coded model with French-consistent as the baseline showed that constructions embedded in French-consistent sentences had significantly longer gaze durations than those embedded in English intact sentences ( $\beta = -0.18$ ,  $SE = 0.02$ ,  $t = -6.60$ ,  $p < 0.01$ ). However, there was no significant difference between French-consistent and French-inconsistent constructions ( $\beta = -0.02$ ,  $SE = 0.02$ ,  $t = -0.98$ ,  $p = 0.32$ ).<sup>7</sup>

<sup>7</sup>A refit model with the reference level set at French-inconsistent showed that first pass gaze durations for French-inconsistent constructions were longer than those embedded in English intact constructions ( $\beta = -0.16$ ,  $SE = 0.02$ ,  $t = -5.60$ ,  $p < 0.01$ ).

For total reading time, a treatment coded model with French-consistent as the baseline showed that constructions embedded in French-consistent sentences were read significantly slower than those embedded in English intact sentences ( $\beta = -0.33$ ,  $SE = 0.03$ ,  $t = -10.08$ ,  $p < 0.01$ ). However, there was no significant difference between French-consistent and French-inconsistent constructions ( $\beta = -0.03$ ,  $SE = 0.03$ ,  $t = -1.20$ ,  $p = 0.22$ ).<sup>8</sup>

**Effects of bilingual language experience.** Similar to the analyses for adjective-noun constructions, we reran the above models including language mixing as an interaction term and general L2 (English) speaking as a control variable, and vice versa. Again, we discuss only significant interaction effects involving condition. Individual differences in general frequency of L2 speaking interacted with the English intact condition for total reading time. Specifically, there was a significant interaction for total reading time between the French-consistent vs. English intact contrast and general frequency of L2 speaking ( $\beta = 0.08$ ,  $SE = 0.03$ ,  $t = 2.67$ ,  $p < 0.01$ ) (see **Figure 3**). This interaction suggests that the more one speaks their L2, the less they experience processing costs for sentences that contain word order violations that are either consistent or inconsistent with French.

Furthermore, there was a significant interaction for total reading time between the French-inconsistent vs. English intact and language mixing ( $\beta = -0.07$ ,  $SE = 0.03$ ,  $t = -2.16$ ,  $p = 0.03$ ) (see **Figure 4**). This interaction suggests that the more one is exposed to environments where both English and French are used in daily life, the more they experience processing costs for sentences that contain word order violations.

<sup>8</sup>A refit model with the reference level set at French-inconsistent showed that total reading time for French-inconsistent constructions were longer than those embedded in English intact constructions ( $\beta = -0.29$ ,  $SE = 0.03$ ,  $t = -8.88$ ,  $p < 0.01$ ).

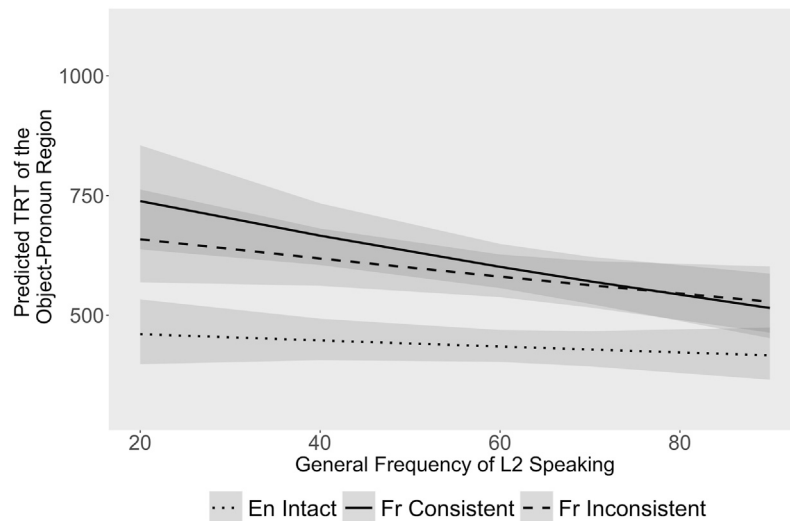


**FIGURE 3 |** Experiment 1 (French-English bilinguals reading in L2), interaction effect of general frequency of L2 speaking (out of 100%) on predicted total reading time (in ms) of the object-pronoun region. Shaded area represents plus/minus one standard error of the mean.

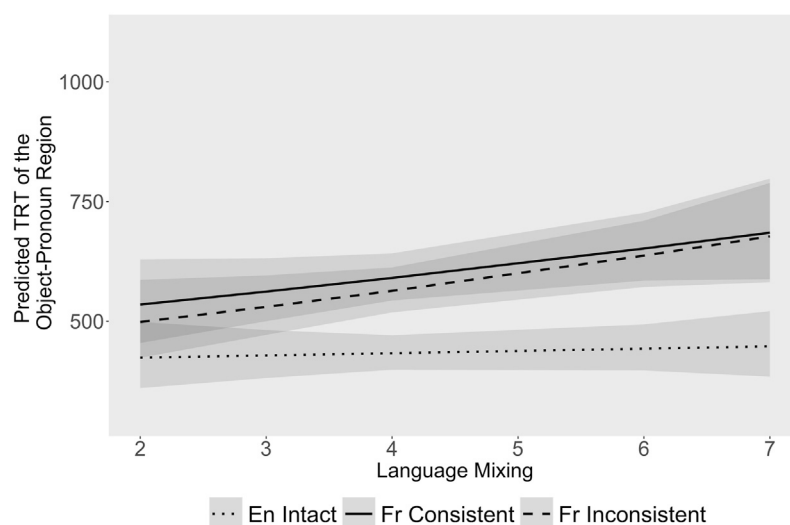


**FIGURE 4 |** Experiment 1 (French-English bilinguals reading in L2), interaction effect of language mixing (out of 7) on predicted total reading time (in ms) of the object-pronoun region. Shaded area represents plus/minus one standard error of the mean.

## Discussion

The results of Experiment 1 were consistent with the idea that bilinguals activated L1 (French) adjective-noun word order when reading in their L2 (English) during early stages of sentence processing. Specifically, when bilinguals read adjective-noun constructions that were ungrammatical in English but felicitous in French, reading times were comparable to intact English sentences, and faster than sentences containing French-inconsistent violations, which may index cross-language syntactic activation. Further, this difference for adjective-noun constructions was reduced for later reading measures,

suggesting that early cross-language activation occurred and was later resolved.

In contrast, object-pronoun constructions showed a different pattern. Here, gaze durations and total reading times for both French-consistent and -inconsistent constructions were longer than gaze durations and total reading times for English intact constructions. Thus, when reading object-pronoun constructions, French-English bilingual participants reading in their L2 showed less tolerance of violations overall. Taken together, the L2 readers tested here appeared to show early syntactic cross-language activation of adjective-noun

constructions that was attenuated for later reading measures, but they never showed cross-language activation of object-pronoun constructions for any reading measure.

Interestingly, individual differences in the general frequency of L2 English speaking and language mixing modulated reading behavior for both constructions. For Adjective-Noun constructions, as can be seen in **Figure 1**, the interaction on early reading measures was driven by the finding that bilingual readers with low general frequency of L2 speaking experience did not show evidence of discriminating between the three conditions at this stage of processing. However, as the general frequency of L2 speaking increased, both intact and French-consistent sentences were read more quickly than the French-inconsistent sentences. This suggests that L2 readers who had less experience with English were less able to distinguish between grammatically correct vs. incorrect sentences, whereas readers with more L2 English experience had comparable gaze durations for both intact and French-consistent sentences that were shorter than for French-inconsistent sentences. Thus, compared to readers with low general frequency of English L2 speaking, they were better able to differentiate intact sentences from sentences that are completely wrong early on.

Moreover, individual differences in general frequency of L2 speaking also played a role during late stages of processing object-pronoun constructions. Total reading times for French-English bilinguals who spoke their L2 frequently suggested greater tolerance of manipulated sentences. In contrast, French-English bilinguals with low frequency of L2 speaking showed a larger preference for the intact condition compared to both other conditions in overall total reading time. This pattern of results may be explained by an association of increased L2 exposure with greater reading fluency, with more fluent readers being able to covertly repair the word order issues in the moment. These readers were quicker at integrating constructions featuring violations than those with lower frequency of English L2 speaking—regardless of consistency with L1 French. Thus, across both constructions, readers who had greater L2 experience generally repaired grammatical violations more easily.

Individual differences in language mixing also modulated later measures of reading for both adjective-noun and object-pronoun constructions. Specifically, for adjective-noun constructions, higher exposure to bilingual discourse was associated with French-consistent violations being processed more similarly to intact sentences. In contrast, for object-pronoun constructions, the same was associated with French-consistent violations being processed more similarly to French-inconsistent constructions. The findings for French-English bilinguals reading in their L2 have implications for sentence processing models, as described in the introduction. Recall that the UCM would predict cross-language activation for both adjective-noun and object-pronoun constructions, because both constructions conflict across languages. Thus, the pattern found for adjective-noun constructions in Experiment 1 (L2 reading), but not the one found for object-pronoun construction, was consistent with the predictions of the UCM.

While the findings of Experiment 1 (L2 reading) suggest that cross-language activation occurs to some extent when bilinguals read sentences containing violations of adjective-noun structures,

it is unclear whether the same pattern of results would occur when bilinguals read in their L1. Some past research has suggested that cross-language activation at a lexical level is less likely to occur during L1 sentence reading (Schwartz and Kroll, 2006; Libben and Titone, 2009; Titone et al., 2011). However, a recent study by Gullifer and Titone (2019) found that under certain circumstances, cross-language activation may in fact be greater for L1 reading because of reduced vigilance with respect to bilingual language control compared to L2 reading. To further investigate these issues within the context of the current experimental design, we investigated the L1 (English) reading patterns of English-French bilinguals, using the same materials and procedures as in Experiment 1.

## EXPERIMENT 2: ENGLISH-FRENCH BILINGUALS READING IN ENGLISH (L1 READING)

### Method

**Participants.** We tested 21 English-French bilinguals (3 men, 18 women) at McGill University and the University of Western Ontario, who did not speak additional languages fluently according to self-report. Participants had a mean age of 21.18 years ( $SD = 4.12$ ), with no uncorrected vision, speech, or reading impairments. Participant recruitment and compensation were as in Experiment 1. Participants reported speaking their L2, French, 20.22% ( $SD = 13.97$ ; Min = 0; Max = 50) of the time. On average, they rated their likelihood to mix French and English as 3.28 out of 7 ( $SD = 1.03$ ; Min = 1; Max = 5).

**Materials.** The materials were the same as in Experiment 1.

**Procedure.** The procedure was the same as in Experiment 1.

**Apparatus.** The apparatus was the same as in Experiment 1.

**Data preprocessing and analytic approach.** The data preprocessing was conducted in the same way and the analytic approach was the same as in Experiment 1.

We excluded observations where the region of interest was skipped. In the adjective-noun experiment, this led us to exclude 19 observations from the first pass gaze duration analysis and 13 observations from the total reading time analysis.<sup>9</sup> In the object-pronoun experiment, this led us to exclude 27 observations from the first pass gaze duration analysis and ten observations from the total reading time analysis.<sup>10</sup> No observations exceeded the upper cut-off.

<sup>9</sup>Of the observations excluded from the first pass gaze duration analysis, eight were presented in the English intact condition, nine in the French-consistent and two in the French-inconsistent condition. Of those, seven of the observations presented in the English intact condition, five of the French-consistent and one of the observations presented in the French-inconsistent condition were also excluded from the total reading time analysis.

<sup>10</sup>Of the observations excluded from the first pass gaze duration analysis, four were presented in the English intact condition, nine in the French-consistent and 13 in the French-inconsistent condition. Of those, one of the observations presented in the English intact condition, five of the French-consistent and four of the observations presented in the French-inconsistent condition were also excluded from the total reading time analysis.

## Results

**Sentence Comprehension.** Sentence comprehension accuracy was 91% among English-French bilinguals, indicating that they were attentive during the reading task.

### Adjective-Noun Constructions

**Core models.** A treatment coded model with French-consistent as the baseline showed that gaze durations for French-consistent constructions were significantly longer than those for English intact constructions ( $\beta = -0.08$ ,  $SE = 0.02$ ,  $t = -3.04$ ,  $p < 0.01$ ), but similar to those for French-inconsistent constructions ( $\beta = 0.03$ ,  $SE = 0.02$ ,  $t = 1.10$ ,  $p = 0.27$ ).<sup>11</sup>

A treatment coded model with French-consistent as the baseline showed that total reading times for French-consistent constructions were longer than those for English intact constructions ( $\beta = -0.26$ ,  $SE = 0.03$ ,  $t = -8.45$ ,  $p < 0.01$ ), but shorter than those for French-inconsistent constructions ( $\beta = 0.11$ ,  $SE = 0.03$ ,  $t = 3.77$ ,  $p < 0.01$ ).<sup>12</sup>

**Effects of bilingual language experience.** Similar to Experiment 1, we reran the above models including language mixing as an interaction term and general frequency of L2 (French) speaking as a control variable, and vice versa. Below, we discuss only significant interaction effects if they involved condition. There was no significant interaction involving language mixing. There was no significant interaction for gaze duration involving general frequency of L2 speaking. There was a significant interaction for total reading time between the French-consistent vs. English intact contrast and L2 speaking ( $\beta = 0.09$ ,  $SE = 0.03$ ,  $t = 2.83$ ,  $p < 0.01$ ), and between the French-inconsistent vs. English intact contrast ( $\beta = 0.07$ ,  $SE = 0.03$ ,  $t = 2.43$ ,  $p = 0.01$ ) (see **Figure 5**). These interactions suggest that the more one speaks their L2 (French), the less they experience processing costs for L1 (English) sentences that contain any word order violations (i.e., whether they are consistent or inconsistent with French).

### Object-Pronoun Constructions

**Core models.** A treatment coded model with French-consistent as the baseline showed that gaze durations for French-consistent constructions were longer than those for English intact constructions ( $\beta = -0.05$ ,  $SE = 0.03$ ,  $t = -1.96$ ,  $p = 0.04$ ), but no different from those for French-inconsistent constructions ( $\beta = -0.01$ ,  $SE = 0.03$ ,  $t = -0.62$ ,  $p = 0.53$ ).<sup>13</sup>

A treatment coded model with French-consistent as the baseline showed that total reading times for French-consistent constructions were longer than those for English intact

constructions ( $\beta = -0.21$ ,  $SE = 0.03$ ,  $t = -6.11$ ,  $p < 0.01$ ), but shorter than those for French-inconsistent constructions ( $\beta = 0.08$ ,  $SE = 0.03$ ,  $t = 2.47$ ,  $p = 0.01$ ).<sup>14</sup>

**Effects of bilingual language experience.** We reran the models above including language mixing as an interaction term and general frequency of L2 speaking as a control variable, and vice versa. There was no significant interaction involving language mixing. Individual differences in general frequency of L2 speaking interacted with the English intact condition for gaze duration. Specifically, there was a significant interaction for gaze duration between the French-consistent vs. English intact contrast and general frequency of L2 speaking ( $\beta = -0.05$ ,  $SE = 0.03$ ,  $t = 1.98$ ,  $p = 0.04$ ) (see **Figure 6**). This interaction suggests that the more English-French bilinguals speak their L2, the more they experience processing costs for L1 sentences that contain word order violations that are consistent with French.

## Discussion

English-French bilinguals reading in their L1 (English) showed limited evidence of cross-language syntactic activation for both construction types. For both adjective-noun and object-pronoun constructions, there was no significant difference between French-consistent and French-inconsistent violations in early measures, whereas for late reading measures, French-consistent violations were read more quickly than French-inconsistent violations, and more slowly than English intact sentences. This suggests that English-French bilinguals were initially sensitive to any violation. However, they were faster overall at integrating both adjective-noun and object-pronoun constructions when they included a violation consistent with their L2 (French) than when the violation was inconsistent with their L2.

Furthermore, individual differences in general frequency of L2 speaking modulated later measures of reading for adjective-noun, whereas it was only associated with early sensitivity to violations of object-pronoun constructions. Total reading times for English-French bilinguals who spoke their L2 frequently exhibited less of a difference between manipulated sentences containing adjective-noun constructions. As can be seen in **Figure 5**, L1 readers who used their L2 about half of the time were quicker at integrating constructions featuring adjective-noun violations than those who spoke their L2 less frequently—regardless of consistency with French.

Of note, the range of general frequency of L2 speaking was greater for the group of French-English bilinguals tested in Experiment 1 who spoke English between 20 and 90% of the time, whereas the English-French bilinguals tested in Experiment 2 spoke French between 0 and 50% of the time. This suggests that English-French bilinguals on the higher end of this spectrum might be balanced bilinguals, whereas French-English on the higher end of the spectrum may be reverse

<sup>11</sup>A refit model with the reference level set at French-inconsistent showed that first pass gaze durations for French-inconsistent constructions were longer than those English intact constructions ( $\beta = -0.11$ ,  $SE = 0.02$ ,  $t = -4.16$ ,  $p < 0.01$ ).

<sup>12</sup>A refit model with the reference level set at French-inconsistent showed that total reading times for French-inconsistent constructions were longer than those in English intact constructions ( $\beta = -0.38$ ,  $SE = 0.03$ ,  $t = -12.24$ ,  $p < 0.01$ ).

<sup>13</sup>A refit model with the reference level set at French-inconsistent showed no significant differences in first pass gaze duration among the sentence types ( $t < 1.96$ ,  $p > 0.05$ ).

<sup>14</sup>A refit model with the reference level set at French-inconsistent showed that total reading times for French-inconsistent constructions were longer than those in English intact constructions ( $\beta = -0.30$ ,  $SE = 0.03$ ,  $t = -8.59$ ,  $p < 0.01$ ).

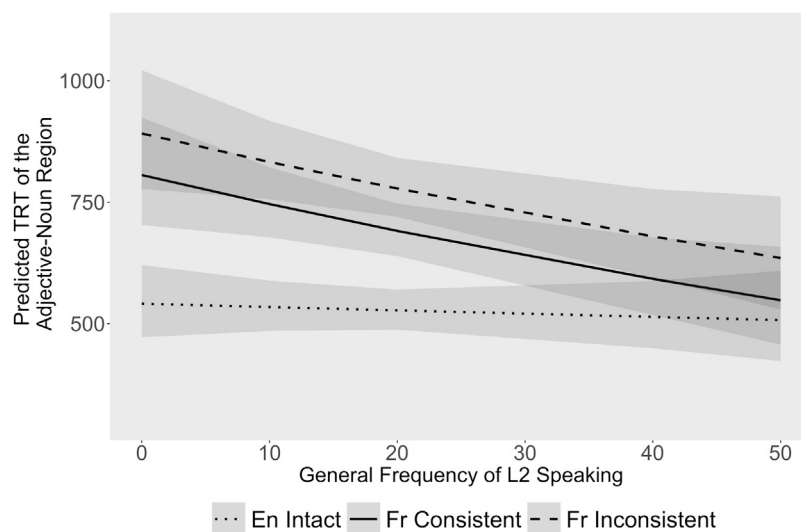


**FIGURE 5 |** Experiment 2 (English-French bilinguals reading in L1), interaction effect of general frequency of L2 speaking (out of 100%) on predicted total reading time (in ms) of the adjective-noun region. Shaded area represents plus/minus one standard error of the mean.

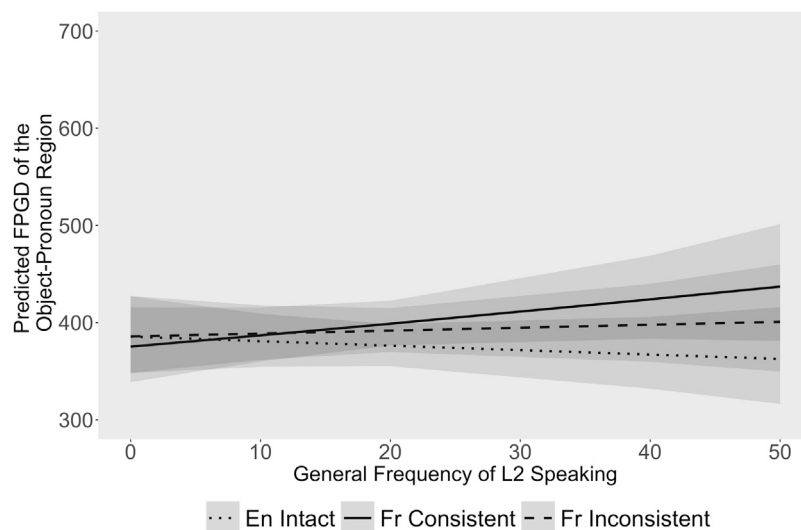


**FIGURE 6 |** Experiment 2 (English-French bilinguals reading in L1), interaction effect of general frequency of L2 speaking (out of 100%) on predicted first pass gaze duration (in ms) of the object-pronoun region. Shaded area represents plus/minus one standard error of the mean.

dominant. This may help explain the different patterns of results found across Experiments 1 and 2. More specifically, the pattern of results observed here may be explained by an association of increased L2 exposure with greater reading fluency and ability to repair violations, whether these stems from intrusion from the L2 or not.

Taken together, the L1 reading data suggested some degree of cross-language syntactic activation for both adjective-noun and object-pronoun constructions during later stages of processing. An extension of the UCM to highly proficient bilinguals predicted that all bilingual readers would exhibit cross-language activation during L1 reading when the

constructions are unshared across the two known languages. However, crucial to the interpretations generated here, tolerance of a violation (i.e., faster reading times) may not only be a result of cross-language influence, but also the number, strength, and position of cues to ungrammaticality. Thus, in Experiment 3, we tested English monolingual reading for the same sentences. To the extent that the above-described results are due to cross-language syntactic activation, people who are functionally monolingual (but may be ambiently exposed to French) should differ to a much lesser degree in their processing of French-consistent and French-inconsistent violations.

## EXPERIMENT 3: FUNCTIONALLY MONOLINGUAL ADULTS READING IN ENGLISH

### Method

**Participants.** We tested 26 functionally English monolingual adults (8 men, 18 women) at the University of Western Ontario. Participants had a mean age of 20.27 years ( $SD = 3.52$ ), with no uncorrected vision, speech, or reading impairments. Participant recruitment and compensation were as in Experiments 1 and 2. Participants had little to no French proficiency (or other languages) based on self-report on the language history questionnaire (Marian et al., 2007), as well as performance on a semantic decision task on French and English words (i.e., animate vs. inanimate judgment; Segalowitz et al., 1995).

**Materials.** The materials were the same as in Experiments 1 and 2.

**Procedure.** The procedure was the same as in Experiments 1 and 2.

**Apparatus.** The apparatus was the same as in Experiments 1 and 2.

**Data preprocessing and analytic approach.** The data preprocessing was conducted in the same way and the analytic approach was the same as in Experiments 1 and 2.

We excluded observations where the region of interest was skipped. In the adjective-noun experiment, this led us to exclude 17 observations from the first pass gaze duration analysis and two observations from the total reading time analysis.<sup>15</sup> In the object-pronoun experiment, this led us to exclude 40 observations from the first pass gaze duration analysis and 27 observations from the total reading time analysis.<sup>16</sup> No observations exceeded the upper cut-off.

### Results

**Sentence Comprehension.** Sentence comprehension accuracy was 90% among English monolingual participants, indicating that they were attentive during the reading task.

#### Adjective-Noun Constructions

**Core models.** A treatment coded model with French-consistent as the baseline showed that constructions embedded in French-consistent sentences had significantly shorter gaze durations than those embedded in French-inconsistent sentences ( $\beta = 0.05$ ,  $SE = 0.02$ ,  $t = 2.12$ ,  $p = 0.03$ ). However, there was only a marginally significant difference between English intact and French-

consistent constructions ( $\beta = -0.04$ ,  $SE = 0.02$ ,  $t = -1.78$ ,  $p = 0.07$ ).<sup>17</sup>

For total reading time, a treatment coded model with French-consistent as the baseline showed that total reading times for constructions embedded in French-consistent sentences were longer than that for English intact sentences ( $\beta = -0.16$ ,  $SE = 0.02$ ,  $t = -5.89$ ,  $p < 0.01$ ) but shorter for French-inconsistent sentences ( $\beta = 0.13$ ,  $SE = 0.02$ ,  $t = 4.73$ ,  $p < 0.01$ ).<sup>18</sup>

#### Object-Pronoun Constructions

**Core models.** A treatment coded model with French-consistent as the baseline showed that gaze durations for French-consistent constructions were longer than those for English intact constructions ( $\beta = -0.08$ ,  $SE = 0.02$ ,  $t = -2.99$ ,  $p < 0.01$ ), but not different from those for French-inconsistent constructions ( $\beta = 0.0$ ,  $SE = 0.02$ ,  $t = 0.89$ ,  $p = 0.37$ ).<sup>19</sup>

A treatment coded model with French-consistent as the baseline showed that total reading times for French-consistent constructions were longer than those for English intact constructions ( $\beta = -0.19$ ,  $SE = 0.03$ ,  $t = -5.90$ ,  $p < 0.01$ ), but only marginally significantly different from French-inconsistent constructions ( $\beta = 0.60$ ,  $SE = 0.03$ ,  $t = 1.81$ ,  $p = 0.06$ ).<sup>20</sup>

### Discussion

The results of Experiment 3 suggest that English speakers who were functionally monolingual were nevertheless sensitive to syntactic violations overall but differed in how sensitive they were to violations of adjective-noun and object-pronoun word order. Overall, this group was less sensitive to violations of adjective-noun constructions than object-pronoun constructions (consistent with the bilingual data reported in Experiments 1 and 2). For adjective-noun constructions, early reading measures suggested instant sensitivity to violations that were inconsistent with French, but not to those consistent with French. Later measures suggested increased sensitivity to both types of violations. For object-pronoun constructions, self-described monolingual readers were consistently sensitive to any type of violation. In sum, these effects suggest that while monolinguals are sensitive to syntactic violations overall, the time course of sensitivity to violations of adjective-noun and object-pronoun constructions differs, independently of any substantial influence of cross-language activation.

<sup>15</sup>Of the observations excluded from the FPGD analysis, six were presented in the English intact condition, six in the French-consistent and five in the French-inconsistent condition. Of those, one of the observations presented in the English intact condition and one of the observations presented in the French-inconsistent condition were also excluded from the TRT analysis.

<sup>16</sup>Of the observations excluded from the FPGD analysis, fourteen were presented in the English intact condition, thirteen in the French-consistent and 13 in the French-inconsistent condition. Of those, ten of the observations presented in the English intact condition, eight of the French-consistent and nine of the observations presented in the French-inconsistent condition were also excluded from the TRT analysis.

<sup>17</sup>A refit model with the reference level set at French-inconsistent showed first pass gaze durations for French-inconsistent constructions were longer than those embedded in English intact constructions ( $\beta = -0.10$ ,  $SE = 0.02$ ,  $t = -3.91$ ,  $p < 0.01$ ).

<sup>18</sup>A refit model with the reference level set at French-inconsistent showed that total reading time for French-inconsistent constructions were longer than those in English intact constructions ( $\beta = -0.30$ ,  $SE = 0.02$ ,  $t = -10.63$ ,  $p < 0.01$ ).

<sup>19</sup>A refit model with the reference level set at French-inconsistent showed that first pass gaze durations for French-inconsistent constructions were longer than English intact constructions ( $\beta = -0.11$ ,  $SE = 0.02$ ,  $t = -3.89$ ,  $p < 0.01$ ).

<sup>20</sup>A refit model with the reference level set at French-inconsistent showed that total reading times for French-inconsistent constructions were longer than those in English intact constructions ( $\beta = -0.25$ ,  $SE = 0.03$ ,  $t = -7.71$ ,  $p < 0.01$ ).

The finding that reading patterns for functional monolingual readers were similar to those exhibited by L2 readers (Experiment 1), clouds our prior interpretation of cross-language activation in Experiments 1 and 2, and suggests that those differences may in fact be a result of certain violations being easier to process than others (i.e., “the car red” is easier to process than “red the car”) because the former is more akin to a simple word transposition error than the latter. These similarities may also stem from the presence of cues to ungrammaticality and their position in the sentence (i.e., “red the car” becomes ungrammatical at “the”, which is earlier than “the car red”, which becomes ungrammatical at “red”). To explore this possibility, we conducted supplementary analyses.

**Supplementary Analyses.** Because readers environmentally exposed to English and French but functionally monolingual exhibited a pattern of results consistent with bilingual adults (but potentially arising from the increased likelihood of word transposition errors for particular items), we conducted an additional analysis to verify the effect of the experimental manipulation for the data in Experiments 1 and 2 (respectively, L2 and L1 reading). Specifically, we verified that the observed effects were not merely due to between-item differences in the strength of the syntactic violation. It is possible that for some items, particular violations were more salient than others, which would be likely to increase reading times for those violations. In order to assess this, we calculated a difference score for each item by subtracting monolinguals’ first pass gaze durations of French-consistent constructions from monolinguals’ first pass gaze durations of English intact constructions. If this yielded a negative difference (i.e., English intact first pass gaze durations < French-consistent first pass gaze durations), we would conclude that this particular item was not susceptible to word transposition errors (i.e., the French-consistent construction was harder to process than the English intact construction).

After doing this, we reran the analyses from Experiment 1 and 2 involving the bilingual samples, controlling for this item-level difference score in the base models. Crucially, the pattern of results was comparable to the original ones, suggesting that they were due to the experimental manipulation. Model comparisons also showed that controlling for the difference score did not improve the fit of any models. However, this exercise highlights one of the challenges that must be addressed when studying multiword constructions in this manner.

In addition, closer inspection of items containing an adjective-noun construction also revealed some ambiguous or alternative interpretations of the sentences, which may have contributed to the observed reading patterns (we thank a careful reviewer for raising this point). First, fifteen of the French-consistent sentences could have potentially been grammatically correct under a set of circumstances that were not true of our sentence materials. For example, the sentence “He seized the shirt wrinkled that needed to be ironed” is ungrammatical because the adjective-noun phrase is followed by the conjunction *that*. This means that the sentence only becomes ungrammatical after the region of interest. Indeed, had the sentence featured an adjectival phrase (e.g., “He seized the

shirt wrinkled by the cats that needed to be ironed”), the adjective-noun word order would have been grammatical. Second, 11 of the French-consistent sentences could have been grammatically correct, assuming that the sentence ended immediately following the region of interest, which was also not true of our sentence materials. For example, the sentence “The father left the house empty that needed to be cleaned” would be grammatical if it had read only “The father left the house empty”, as it may be interpreted as a resultative construction.

To ensure that this serendipitous variability across items did not systematically impact the results, we took the following steps. First, we coded the items to be “as intended”, “not as intended”. We then reran our analysis on only the 34 items labeled “as intended”, while still controlling for the above-described item-level difference score (i.e., subtracting monolinguals’ first pass gaze durations of French-consistent constructions from monolinguals’ gaze durations of English intact constructions). Crucially, this did not substantially change the results, suggesting that the patterns of results observed in Experiments 1 and 2 were not likely driven by these problematic items. A summary of the models used in both supplementary analyses can be found in the OSF repository ([https://osf.io/jec5s/?view\\_only=a9d0ad4f9b994cd9b93371d2a2089cf1](https://osf.io/jec5s/?view_only=a9d0ad4f9b994cd9b93371d2a2089cf1)).

## GENERAL DISCUSSION

In this study, we addressed the following questions: 1) Do bilingual readers show evidence of cross-language syntactic activation during L2 and L1 reading of adjective-noun and object-pronoun constructions (Experiments 1 and 2); 2) Do individual differences in language mixing and general frequency of L2 speaking modulate cross-language activation patterns during bilingual reading (Experiments 1 and 2); and 3) How do monolingual reading patterns of adjective-noun and object-pronoun constructions compare to those of bilinguals (Experiment 3). We discuss each in turn.

- 1) Do bilingual readers show evidence of cross-language syntactic activation during L2 and L1 reading of adjective-noun and object-pronoun constructions (Experiments 1 and 2)

In L2 reading (Experiment 1), we found that French-English bilinguals reading sentences that contained adjective-noun constructions were less sensitive to the French-consistent violation during early stage reading, as evidenced by comparable gaze durations for English intact and French-consistent constructions. Sensitivity increased during later stages of reading, where French-consistent constructions were “in the middle”; they were slower than English intact constructions, but faster than French-inconsistent constructions. This pattern suggests that L1 French adjective-noun syntactic frame was active during early stages of L2 English processing, and to a lesser extent, during late stages. When the same participants read sentences containing object-pronoun constructions, a French-consistent violation did not behave the same way. Specifically, there was no difference in how French-

English bilinguals read French-consistent vs. inconsistent violations, suggesting that no cross-language activation of the L1 French object-pronoun construction occurred during L2 English processing.

In L1 reading (Experiment 2), English-French bilinguals were initially sensitive to all violations, whether these were consistent with their L2 (French) or not. For both adjective-noun and object-pronoun constructions, there was no difference between French-consistent and French-inconsistent violations in early measures. During later stages of processing, L1 readers were somewhat less sensitive to violations that were consistent with their L2 (French), reading them more quickly than French-inconsistent violations. Overall, this pattern suggests some degree of activation of the L2 syntactic frames during late stages of L1 processing.

Existing models of bilingual sentence processing, such as the UCM, predict reduced sensitivity to syntactic violations that are consistent with the other known language as a result of cross-language activation (to the extent that the constructions conflict across languages). Moreover, a strict interpretation of the UCM would predict even greater tolerance for violations of the object-pronoun construction, as these constructions conflict across English and French to a greater degree than adjective-noun constructions. Here, this would have resulted in faster reading times for French-consistent compared to French-inconsistent violations, for adjective-noun constructions, and even more so for object-pronoun constructions. However, our results only partially align with these predictions as increased tolerance for French-consistent vs. French-inconsistent violations was only observed for adjective-noun constructions in both experiments, but not for object-pronoun constructions. Interestingly, tolerance for French-consistent violations was more apparent at early stages of processing for French L1 bilinguals, while it was present only during late stages of processing for English L1 bilinguals. This finding suggests that to the extent that cross-language activation of non-target constructions occurs, it may impact different aspects of processing—initial lexical processing of the construction for L2 readers, ambiguity resolution/error repair processes for L1 readers.

We emphasize here that the UCM formulates predictions based on offline behaviors (i.e., grammaticality judgements) and does not make explicit predictions about online behavior (i.e., eye movements of reading). Here, rather than looking at the impact of cross-language activation on an explicit grammaticality judgement, we investigated the time course of cross-language activation during natural reading in the absence of explicit grammaticality judgements. Thus, in applying this model here, we are assuming that comprehension outcome and process are related to some degree. Additionally, the UCM model makes no specific predictions regarding the dynamic manipulation of cues during processing, which makes it difficult to formulate clear predictions about online sensitivity during various stages of processing, and thus may explain a lack of alignment of our results with model predictions.

- 2) Do individual differences in language mixing and general frequency of L2 speaking modulate cross-language activation patterns during bilingual reading (Experiments 1 and 2)

In addition to examining reading time differences across the sentence conditions at the group level, we also examined whether individual differences in L2 experience among readers modulated cross-language activation. When French-English bilinguals (L2 reading) read adjective-noun constructions, more frequent English L2 speaking was associated with increased tolerance to the French-consistent violation on early measures of reading, whereas for object-pronoun constructions, the same was associated with higher tolerance to violations overall on late measures of reading. Among the same participants, late measures of reading showed that higher language mixing was associated with more similar processing of English intact and French-consistent adjective-noun constructions, whereas the same was associated with reduced overall tolerance to any type of violation of object-pronoun constructions. When English-French bilinguals (L1 reading) read adjective-noun constructions, late measures showed that higher frequency of French L2 speaking was associated with increased tolerance to either type of violation. When the same participants read object-pronoun constructions, early measures showed that higher frequency of French L2 speaking was associated with somewhat decreased tolerance to violations consistent with French.

In the literature, there is some empirical evidence suggesting that syntactic structures for each language are represented separately for bilinguals at low levels of proficiency, whereas syntactic representations may become increasingly shared for proficient bilinguals, leading to greater cross-language syntactic activation (Bernolet et al., 2009; reviewed in; Hartsuiker and Pickering, 2008). This may explain why French-English bilinguals with high frequency of L2 English usage were more tolerant of French-consistent violations of the adjective-noun constructions.<sup>21</sup> Moreover, as we also found increased tolerance to these violations for French-English bilinguals with high frequency of language mixing, language mixing may also contribute to the integration of syntactic representations across languages. In contrast, as high frequency of L2 French speaking was associated with increased tolerance of all types of violations of the adjective-noun in English-French bilinguals, it may suggest that high levels of experience with L2 may enhance flexibility in processing syntactic violations during L1 reading independently of whether these are consistent with L2.

In contrast with adjective-noun constructions, the individual differences pattern for object-pronoun constructions suggests that not all constructions have a comparable status in the bilingual mind as bilingual experience never uniquely impacted the processing of French-consistent violations of object-pronoun constructions, it is possible that these constructions, which conflict maximally in word order across English and French, are stored separately (see Loebell and Bock, 2003). In contrast, constructions with word order overlap may become increasingly shared as bilingual experience increases

<sup>21</sup>It should be noted, however, that these results contradict earlier findings suggesting that age of acquisition is more critical for syntactic representation than proficiency (e.g., Wartenberger et al., 2003).

(Hartsuiker et al., 2004; Bernolet et al., 2009). We discuss other differences between the two constructions that may have contributed to this effect below.

### 3) How do monolingual reading patterns of adjective-noun and object-pronoun constructions compare to those of bilinguals (Experiment 3)

Functionally monolingual readers were somewhat more tolerant of French-consistent violations to adjective-noun word order than they were of French-inconsistent violations, especially during early reading measures. However, no such tolerance was observed for violations of object-pronoun word order. For object-pronoun word order, functional monolinguals read both French-consistent and French-inconsistent constructions more slowly than English intact constructions, across early and late reading measures.

The observed reading patterns suggest that while monolinguals are sensitive to syntactic violations overall, the time course of sensitivity to violations of adjective-noun and object-pronoun constructions differs, independently of any substantial influence of cross-language activation. The difference in processing patterns across the two types of syntactic constructions may be a result of the cues to ungrammaticality that each construction provides. Thus, it is possible that the violation to adjective-noun word order was noticeable but did not impact readers' ability to comprehend the sentence, as adjectives are not crucial to sentence grammaticality. In contrast, the violation may have been more harmful to the comprehension of sentences containing an object-pronoun construction, as omission of object-pronouns in transitive clauses directly causes ungrammaticality.

While monolingual reading patterns cannot be influenced by cross-language activation, we did find that the reading patterns of functional English monolinguals resembled those exhibited by French-English bilinguals reading in their L2 (Experiment 1). Specifically, both functionally monolingual and L2 readers showed initial tolerance of the French-consistent violation that was attenuated during later stages of processing, whereas English-French bilinguals reading in their L1 showed no tolerance on early measures of processing and only limited tolerance of the French-consistent violations on late stages of processing.

It is important to acknowledge several limitations in the selection of stimuli in the adjective-noun experiment that may have contributed to the first pass gaze duration patterns of monolingual and L2 readers. Specifically, some of the French-consistent sentences included in our experiment do not become ungrammatical until after the region of interest. This could have led to the increased tolerance of the French-consistent violation during the early stages of reading. However, the results of our supplemental analysis reproduced the above-described effect with a subset of only those items that included an unambiguous violation, this does not appear to be driving the observed pattern of effects. Nevertheless, it remains possible that idiosyncratic aspects of the particular sentences we used could have impacted the results in some fashion. Thus, future work should more carefully consider the

interaction of grammatical structure and lexical choice with respect to these issues.

Perhaps, the observed pattern of results could be explained by the common observation that for functionally monolingual readers, processing of frequent constructions, such as adjective-noun and object-pronoun constructions, may be automatic and effortless. There is also evidence that monolinguals have a larger perceptual span or area of effective visual perception during reading, which means that they likely make fewer fixations to comprehend a phrase (Whitford and Titone, 2012; Whitford and Titone 2015). Both of these factors may lead to higher tolerance of syntactic violations, because their efficient processing capacity facilitates recovery from errors. L2 readers show similar patterns, though potentially for a different reason. Their reduced proficiency/relative exposure to the L2 may result in a smaller perceptual span (Whitford and Titone, 2015) and less automatized L2 syntactic processing compared to L1 (Favreau and Segalowitz, 1983). This, combined with non-selective activation, is likely to impact L2 processing, resulting ultimately in higher tolerance of violations. Lastly, L1 reading (Experiment 2) may also be affected by non-selective activation, in that increased control is required to suppress transfer from the L2 (see Gullifer and Titone, 2019), while also benefitting from stronger activation of the L1 syntactic frames and a perceptual span similar to the one exhibited by monolinguals (Whitford and Titone, 2012; Whitford and Titone 2015). Ultimately, this may result in decreased tolerance to violations when reading in L1.

As this account suggests, it is possible that the reading patterns observed across all three experiments were not exclusively a result of the presence or absence of cross-language activation, but rather a result of an interplay of factors, including word transposition characteristics as well as cues to ungrammaticality, which are likely correlated. For example, a recent study found that French monolingual participants were slow and generally inaccurate when judging the grammaticality of sentences with transposed words (Mirault et al., 2018). Importantly, these transposition effects were more salient for function words (e.g., pronouns) compared to content words (e.g., adjectives and nouns). The researchers suggest that this pattern is linked to the role each type of word plays in making up a grammatical sentence. Specifically, function words allow the reader to quickly build a syntactic frame, which is then filled in with content words. Thus, a violation of the structure of the frame (i.e., pronoun word-order violation) may be more detrimental to perceived grammaticality than a violation of the content word position. This interpretation is in line with what we have found for both monolingual and bilingual reading in that across groups, tolerance of violations was generally lower for object-pronoun constructions than adjective-noun constructions.

Given this account, the processing pattern we observed in this study point to the presence of word transposition effects. Some languages, such as German, which rely heavily on a case system (Bornkessel and Schlesewsky, 2006), are highly flexible in their word order and as a result, their speakers may be less sensitive to word order transpositions. In both English and French, however, word order is crucial to comprehension. Thus, the overall sensitivity to

word transpositions in this experiment should be comparable across the types of constructions and groups of readers, hence it is critical that a monolingual control group be included. There may be less transfer between language systems when word order is a reliable indicator of grammatical function across the L1 and L2. Indeed, our pattern of results suggests that any possible effects of cross-language activation may be diminished in the presence of a cue as strong as word order in English and French. Thus, when investigating cross-linguistic effects, it is crucial to consider strength of grammatical cues across language pairs (see Mirault et al., 2018, for a similar argument).

Another intriguing possibility is that in multilingual communities where particular languages routinely converge and mix (e.g., English and French in Montreal), the ambient linguistic environment is more likely to contain language-mixed constructions (i.e., syntactic “accents”) produced by people who are prone to making such errors. For example, in Montreal a frequent example of such a syntactic accent is when English speakers say, “close the lights” vs. “shut the lights”. What makes this example compelling is that “close the lights” is not normative for English, however, because it is the English translation equivalent of the normative French construction “fermer la lumière”, French-English bilinguals routinely produce that utterance. In turn, because that utterance finds itself in the linguistic environment, that makes it more frequently encountered for all language users, including language users who are functionally monolingual. Indeed, recent work by Bice and Kroll (2019), compared monolinguals who live in a linguistically diverse context (i.e., California) with monolinguals living in a linguistically homogenous context (i.e., Pennsylvania) on their abilities to learn vowel harmony in Finnish. Their findings suggest that exposure to linguistic diversity promotes new language learning. In other words, there can be “bilingual effects” on functional monolinguals within a geographic multilingual community.

In fact, bilinguals’ tolerance of cross-language patterns may be similarly affected by linguistic diversity in their environment. As such, tolerance to violations might not be due to something internal to bilinguals’ language representation, but rather a result of the input they receive from other bilinguals in their community that includes these wrong patterns. Thus, in a usage-based fashion, some French-consistent patterns in English might become more acceptable over time (e.g., close the lights in Quebec) leading to the development of “syntactic accents”. Importantly, to the extent that such effects can occur, it would lead to complexities in perfectly isolating pure “cross-language activation” of any type of construction. Thus, while we cannot directly address this issue currently in this particular study, we are mindful that such effects may be at play in a manner that is highly worthy of future investigation.

Lastly, it is important to keep in mind that most studies have employed offline tasks (such as grammaticality judgements), rather than online tasks (such as eye tracking). Thus, future studies may benefit from an approach combining these two measures to investigate which parts of a construction cause a sentence to be perceived as ungrammatical. As such, fixations on individual words within a construction, and their relationship to accuracy in grammatical judgements, should be investigated. As well, the between-group differences in this study point to a potential role for non-target language control in resolving syntactic violations. Investigating the impact of individual differences in executive

control on bilingual syntactic processing is a desirable avenue for future work.

To conclude, the experiments presented in this manuscript provide some evidence about the presence of cross-language syntactic activation. For both L2 and L1 reading, we observed patterns consistent with cross-language influence for adjective-noun constructions, after item-level variability derived from monolingual readers was taken into account. Moreover, individual differences in L2 experience modulated both L2 and L1 reading for adjective-noun constructions. However, none of these effects emerged when the same participants read object-pronoun constructions. Thus, the totality of evidence leads us to tentatively conclude the presence of cross-language activation for adjective-noun constructions. Nevertheless, we caution future researchers examining this issue to be mindful of the many methodological complexities associated with investigating multiword constructions of this type using the violation method, and specifically attend to the fact that a lot more could be going on in these and other past reports than simply “pure” cross-language syntactic activation.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material. Data and code can be accessed from [https://osf.io/jec5s/?view\\_1807onlya9d0ad4f9b994cd9b93371d2a2089cf1](https://osf.io/jec5s/?view_1807onlya9d0ad4f9b994cd9b93371d2a2089cf1).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the McGill University Research Ethics Board and the University of Western Ontario Ethics Board. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

DT designed the experiment. DT created the items and VW modified them. VW created the eye-tracking scripts to enable data collection. NV, JG, VW, DF, DJ, and DT contributed to the data acquisition. NV, PP, JG, and DT processed, analyzed, and interpreted the data. NV and PP drafted the final version of the manuscript, DT supervised the writing, and the remaining authors reviewed and provided critical feedback. All authors thus made a substantial and direct contribution to the work, and all have approved it for publication.

## FUNDING

This work was supported by an NSERC Discovery Grant and Canada Research Chair Award to DT and an NSERC Discovery Grant to DJ.

## REFERENCES

- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *J. Mem. Lang.* 68 (3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* 67 (1), 1–48. doi:10.18637/jss.v067.i01
- Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2009). Persistence of Emphasis in Language Production: A Cross-Linguistic Approach. *Cognition* 112 (2), 300–317. doi:10.1016/j.cognition.2009.05.013
- Bice, K., and Kroll, J. F. (2019). English Only? Monolinguals in Linguistically Diverse Contexts Have an Edge in Language Learning. *Brain Lang.* 196, 104644. doi:10.1016/j.bandl.2019.104644
- Bornkessel, I., and Schlesewsky, M. (2006). The Extended Argument Dependency Model: A Neurocognitive Approach to Sentence Comprehension across Languages. *Psychol. Rev.* 113 (4), 787–821. doi:10.1037/0033-295X.113.4.787
- Clifton, C., Jr, Staub, A., and Rayner, K. (2007). “Eye Movements in reading Words and Sentences,” in *Eye movements: A window on mind and brain*. Editors R. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill (Oxford, United Kingdom: Elsevier), 341–372. doi:10.1016/b978-008044980-7/50017-3
- Dijkstra, T., and Van Heuven, W. J. B. (2002). The Architecture of the Bilingual Word Recognition System: From Identification to Decision. *Bilingualism* 5 (3), 175–197. doi:10.1017/S1366728902003012
- Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., et al. (2019). Multilink: a Computational Model for Bilingual Word Recognition and Word Translation. *Bilingualism* 22 (4), 657–679. doi:10.1017/S1366728918000287
- Dussias, P. E., and Sagarra, N. (2007). The Effect of Exposure on Syntactic Parsing in Spanish-English Bilinguals. *Bilingualism* 10 (1), 101–116. doi:10.1017/S13667289060002847
- Dussias, P., Dietrich, A. J., and Villegas, Á. (2015). “Cross-language Interactions during Bilingual Sentence Processing,” in *The Cambridge Handbook of Bilingual Processing*. Editor J. Schwieter (Cambridge: Cambridge University Press), 349–366. doi:10.1017/CBO9781107447257.016
- Falkenstein, M., Hoormann, J., and Hohnsbein, J. (1999). ERP Components in Go/Nogo Tasks and Their Relation to Inhibition. *Acta Psychologica* 101 (2–3), 267–291. doi:10.1016/S0001-6918(99)00008-6
- Favreau, M., and Segalowitz, N. S. (1983). Automatic and Controlled Processes in the First- and Second-Language reading of Fluent Bilinguals. *Mem. Cogn.* 11, 565–574. doi:10.3758/bf03198281
- Fox, J., and Weisberg, S. (2018). Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. *J. Stat. Soft.* 87 (9), 1–27. doi:10.18637/jss.v087.i09
- French-Mestre, C. (2005). Eye-movement Recording as a Tool for Studying Syntactic Processing in a Second Language: A Review of Methodologies and Experimental Findings. *Second Lang. Res.* 21 (2), 175–198. doi:10.1191/0267658305sr257oa
- Friesen, D. C., Ward, O., Bohnet, J., Cormier, P., and Jared, D. (2020). Early Activation of Cross-Language Meaning from Phonology during Sentence Processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 46 (9), 1754–1767. doi:10.1037/xlm0000849
- Green, D. W., and Wei, L. (2014). A Control Process Model of Code-Switching. *Lang. Cogn. Neurosci.* 29 (4), 499–511. doi:10.1080/23273798.2014.882515
- Green, D. W., and Wei, L. (2016). Code-switching and Language Control. *Bilingualism* 19 (5), 883–884. doi:10.1017/s1366728916000018
- Grosjean, F. (2001). “The Bilingual’s Language Modes,” in *One Mind, Two Languages: Bilingual Language Processing*. Editor J. Nicol (Oxford, UK: Blackwell), 1–22.
- Gullifer, J. W., and Titone, D. (2019). The Impact of a Momentary Language Switch on Bilingual reading: Intense at the Switch but Merciful Downstream for L2 but Not L1 Readers. *J. Exp. Psychol. Learn. Mem. Cogn.* 45 (11), 2036–2050. doi:10.1037/xlm0000695
- Hartsuiker, R. J., and Pickering, M. J. (2008). Language Integration in Bilingual Sentence Production. *Acta Psychologica* 128 (3), 479–489. doi:10.1016/j.actpsy.2007.08.005
- Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is Syntax Separate or Shared between Languages? *Psychol. Sci.* 15 (6), 409–414. doi:10.1111/j.0956-7976.2004.00693.x
- Hirofani, M., Frazier, L., and Rayner, K. (2006). Punctuation and Intonation Effects on Clause and Sentence Wrap-Up: Evidence from Eye Movements. *J. Mem. Lang.* 54 (3), 425–443. doi:10.1016/j.jml.2005.12.001
- Jared, D. (2015). “Literacy and Literacy Development in Bilinguals,” in *Oxford Library of Psychology. The Oxford Handbook of reading*. Editors A. Pollatsek and R. Treiman (Oxford, UK: University Press), 165–182.
- Kasparian, K., and Steinhauer, K. (2017). When the Second Language Takes the lead: Neurocognitive Processing Changes in the First Language of Adult Attriters. *Front. Psychol.* 08, 389. doi:10.3389/fpsyg.2017.00389
- Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends Cogn. Sci.* 22 (2), 154–169. doi:10.1016/j.tics.2017.11.006
- Kroll, J. F., and Tokowicz, N. (2005). “Models of Bilingual Representation and Processing,” in *Handbook of Bilingualism: Psycholinguistic Approaches*. Editors J. F. Kroll and A. M. B. de Groot (New York, NY: Oxford University Press), 531–553.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Soft.* 82 (13), 1–26. doi:10.18637/jss.v082.i13
- Lauro, J., and Schwartz, A. I. (2017). Bilingual Non-selective Lexical Access in Sentence Contexts: A Meta-Analytic Review. *J. Mem. Lang.* 92, 217–233. doi:10.1016/j.jml.2016.06.010
- Libben, M. R., and Titone, D. A. (2009). Bilingual Lexical Access in Context: Evidence from Eye Movements during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 35 (2), 381–390. doi:10.1037/a0014875
- Loebell, H., and Bock, K. (2003). Structural Priming across Languages. *Linguistics* 41 (5), 791–824. doi:10.1515/ling.2003.026
- MacWhinney, B. (1987). “The Competition Model,” in *Mechanisms of Language Acquisition*. Editor B. MacWhinney (Hillsdale, NJ: Lawrence Erlbaum), 249–308.
- MacWhinney, B. (2005). “A Unified Model of Language Acquisition,” in *Handbook of Bilingualism: Psycholinguistic Approaches*. Editors J. F. Kroll and A. M. B. de Groot (New York, NY: Oxford University Press), 49–67.
- Mahowald, K., and Fedorenko, E. (2016). Reliable Individual-Level Neural Markers of High-Level Language Processing: A Necessary Precursor for Relating Neural Variability to Behavioral and Genetic Variability. *NeuroImage* 139, 74–93. doi:10.1016/j.neuroimage.2016.05.073
- Mirault, J., Snell, J., and Grainger, J. (2018). You that Read Wrong Again! A Transposed-word Effect in Grammaticality Judgments. *Psychol. Sci.* 29 (12), 1922–1929. doi:10.1177/0956797618806296
- Palma, P., and Titone, D. (2020). “Bilingual Lexical Access and reading,” in *Bilingual Lexical Ambiguity Resolution*. Editors R. R. Heredia and A. M. B. Cieslicka (New York, NY: Cambridge University Press), 159–183. doi:10.1017/9781316535967.008
- Pivneva, I., Mercier, J., and Titone, D. (2014). Executive Control Modulates Cross-Language Lexical Activation during L2 reading: Evidence from Eye Movements. *J. Exp. Psychol. Learn. Mem. Cogn.* 40 (3), 787–796. doi:10.1037/a0035583
- Roberts, L. (2012). Individual Differences in Second Language Sentence Processing. *Lang. Learn.* 62, 172–188. doi:10.1111/j.1467-9922.2012.00711.x
- Schwartz, A. I., and Kroll, J. F. (2006). Bilingual Lexical Activation in Sentence Context. *J. Mem. Lang.* 55 (2), 197–212. doi:10.1016/j.jml.2006.03.004
- Segalowitz, N., Watson, V., and Segalowitz, S. (1995). Vocabulary Skill: Single-Case Assessment of Automaticity of Word Recognition in a Timed Lexical Decision Task. *Sec. Lang. Res.* 11 (2), 121–136.
- Snell, J., and Grainger, J. (2019). Word Position Coding in reading Is Noisy. *Psychon. Bull. Rev.* 26 (2), 609–615. doi:10.3758/s13423-019-01574-0
- Street, J. A., and Dąbrowska, E. (2014). Lexically Specific Knowledge and Individual Differences in Adult Native Speakers’ Processing of the English Passive. *Appl. Psycholinguistics* 35 (1), 97–118. doi:10.1017/S0142716412000367
- Thierry, G., and Sanoudaki, E. (2012). Activation syntaxique non-sélective à la langue chez le bilingue précoce. *Revue Française de Linguistique Appliquée* 17 (2), 33–48. doi:10.3917/rfla.172.0033
- Titone, D., Libben, M., Mercier, J., Whitford, V., and Pivneva, I. (2011). Bilingual Lexical Access during L1 Sentence reading: The Effects of L2 Knowledge, Semantic Constraint, and L1-L2 Intermixing. *J. Exp. Psychol. Learn. Mem. Cogn.* 37 (6), 1412–1431. doi:10.1037/a0024492
- Titone, D., Whitford, V., Lijewska, A., and Itzhak, I. (2016). “Chapter 1. Bilingualism, Executive Control, and Eye Movement Measures of reading,” in *Cognitive Control*

- and Consequences of Multilingualism. Editor J. W. Schwieter (Amsterdam/Philadelphia: John Benjamins Publishing Company), 11–46. doi:10.1075/bpa.2.02tit
- Tuninetti, A., Warren, T., and Tokowicz, N. (2015). Cue Strength in Second-Language Processing: An Eye-Tracking Study. *Q. J. Exp. Psychol.* 68 (3), 568–584. doi:10.1080/17470218.2014.961934
- Van Assche, E., Duyck, W., Hartsuiker, R. J., and Diependaele, K. (2009). Does Bilingualism Change Native-Language Reading? *Psychol. Sci.* 20 (8), 923–927. doi:10.1111/j.1467-9280.2009.02389.x
- Assche, E. V., Duyck, W., and Hartsuiker, R. J. (2012). Bilingual Word Recognition in a Sentence Context. *Front. Psychol.* 3, 174. doi:10.3389/fpsyg.2012.00174
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., and Perani, D. (2003). Early Setting of Grammatical Processing in the Bilingual Brain. *Neuron* 37 (1), 159–170. doi:10.1016/S0896-6273(02)01150-9
- Wells, J., Christiansen, M., Race, D., Acheson, D., and MacDonald, M. (2009). Experience and Sentence Processing: Statistical Learning and Relative Clause Comprehension. *Cogn. Psychol.* 58 (2), 250–271. doi:10.1016/j.cogpsych.2008.08.002
- Whitford, V., and Titone, D. (2012). Second-language Experience Modulates First- and Second-Language Word Frequency Effects: Evidence from Eye Movement Measures of Natural Paragraph reading. *Psychon. Bull. Rev.* 19 (1), 73–80. doi:10.3758/s13423-011-0179-5
- Whitford, V., and Titone, D. (2015). Second-language Experience Modulates Eye Movements during First- and Second-Language Sentence reading: Evidence from a Gaze-Contingent Moving Window Paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 41 (4), 1118–1129. doi:10.1037/xlm0000093
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Vingron, Palma, Gullifer, Whitford, Friesen, Jared and Titone. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Cognitive and Linguistic Skills Associated With Cross-Linguistic Transfer in the Production of Oral Narratives in English as a Foreign Language by Arabic- and Hebrew-Speaking Children: Finding Common Denominators

## OPEN ACCESS

### Edited by:

Anat Prior,  
University of Haifa, Israel

### Reviewed by:

Sharon Armon-Lotem,  
Bar-Ilan University, Israel  
Sibylla Leon Guerrero,  
Harvard University, United States

### \*Correspondence:

Susie Russak  
susie.russak@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 04 February 2021

**Accepted:** 06 July 2021

**Published:** 09 August 2021

### Citation:

Russak S and Zaretsky E (2021)  
Cognitive and Linguistic Skills  
Associated With Cross-Linguistic  
Transfer in the Production of Oral  
Narratives in English as a Foreign  
Language by Arabic- and  
Hebrew-Speaking Children: Finding  
Common Denominators.  
Front. Psychol. 12:664152.  
doi: 10.3389/fpsyg.2021.664152

**Susie Russak<sup>1\*†</sup> and Elena Zaretsky<sup>2†</sup>**

<sup>1</sup> English Teacher Training Track, Faculty of Education, Beit Berl College, Kfar Saba, Israel, <sup>2</sup> Department of Psychology, Clark University, Worcester, MA, United States

Many studies have examined literacy and related skills among learners of English as a foreign language (EFL), but little attention has been given to the role of oral language within a cross-linguistic framework despite the fact that English is the most widely spoken additional language today. Oral narratives rely on lexical, morphosyntactic, and conceptual knowledge. An in-depth examination of this modality can shed light on specific associations between cognitive and linguistic L1 and EFL skills and suggest possible mediating variables that assist multilingual speakers in producing complete oral narratives in EFL. The present study examined L1 and EFL contributors to EFL oral narratives produced by native Arabic ( $n = 85$ ) and Hebrew ( $n = 86$ ) speaking sixth graders seeking to identify cross-linguistic influences. We assessed general cognitive skills, phonological memory (PM), lexical, morphosyntactic knowledge, and reading comprehension in L1 (Hebrew speakers), Modern Standard Arabic (MSA, L2), L3 Hebrew (for Arabic speakers) and EFL. The “Cookie Theft” task assessed EFL elicited narratives using modified narrative analysis scales to account for microstructure (lexical and morphosyntactic complexity) and macrostructure (understanding story elements), generating a Total Narrative score. Our results yielded different patterns of underlying psycholinguistic profiles, and cross and within language associations for each group. Strong interactions between L1, L2/L3, and EFL morphological awareness and reading comprehension suggested cross-linguistic transfer. Regression analysis identified the most influential skills supporting EFL narratives for each linguistic group: English reading comprehension (ERC) was essential for Hebrew speakers and English morphological awareness (EMA) for Arabic ones. These results suggested different allocations of cognitive and linguistic resources in EFL narratives. The results also allowed to identify a common mediating skill for both groups. Findings are discussed within the theoretical

framework of the *Interdependence Hypothesis*, the *Linguistic Proximity Model*, as well as accounts of direct and indirect transfer, which illuminate the impact of typological distance, general language proficiency and components of linguistic knowledge on cross-linguistic transfer in EFL oral language production.

**Keywords:** cross-linguistic influence, typological distance, oral narratives, morphosyntactic structures, Semitic languages, English as a foreign language

## INTRODUCTION

In today's world, approximately half of the population is multilingual (Grosjean, 2010), and many children acquire literacy in school in a language that they do not speak at home (Nag et al., 2019). Hence, the need to understand the role of crosslinguistic influence (CLI) in both oral and written language domains has become even more relevant. Within this framework, the nature of cross linguistic influences has been attributed to the typological proximity among the languages, the linguistic repertoire of the learners: how many languages they know (Cenoz, 2013), and the levels of proficiency within each language (*Interdependence Hypothesis*, Cummins, 1979, 2014). While CLIs in the domain of written language have been researched extensively, including crosslinguistic transfer of phonological and morphological skills (Schwartz et al., 2007; Saiegh-Haddad and Geva, 2008; Luo et al., 2014), the domain of oral language remains the “Cinderella” of crosslinguistic enquiries. Yet examining oral language skills can provide a window to the repertoire of resources of multilingual learners (Boerma et al., 2016). Oral narrative skills are reliant on underlying cognitive abilities, such as working memory (Kormos and Trebits, 2011), in addition to strong lexical and syntactic knowledge (Dickinson et al., 2019), and activation of metacognitive skills (Cortazzi and Jin, 2007; Kupersmitt et al., 2014). Moreover, they comprise a rich source of information regarding language development (Gagarina et al., 2015). In light of the prevalence of multilingualism in today's society, an examination of cross linguistic influence in the oral language domain among typologically distant languages is warranted.

The present study examined English as a foreign language (EFL) oral narratives skill among sixth grade native Arabic and Hebrew speaking children. The typological distance between English, a Germanic language, and Hebrew and Arabic, both Semitic languages, is obvious. However, while Arabic and Hebrew share typological characteristics, Arabic is unique in that it is a diglossic language (Leikin et al., 2014; Saiegh-Haddad and Henkin-Roitfarb, 2014). This means that Arabic speaking children, while speaking a vernacular of Arabic at home, are exposed to the Modern Standard form when they begin school as their second spoken and first written language. They then begin to study Hebrew as their second written and third spoken language in the second grade, and in third grade, they begin to study English as their third written language and fourth spoken language. Thus, we were interested in exploring the cross linguistic impact of this specific multilingual setting on EFL oral language production, particularly in light of the fact that all children in Israel study English as their first foreign language according to the same national curriculum.

Of additional importance, we were interested in exploring the possible mediating factors that may support EFL oral language skills in this sample of speakers of two different Semitic languages.

## Cross Linguistic Influences of Cognitive and Linguistic Skills on Oral Language Production

The term “crosslinguistic influence” describes cognitive and linguistic processes that allow to apply knowledge of one language to another language (Moattarian, 2013). Cross linguistic influences in language production among speakers of multiple languages have been attributed to learner characteristics, including number of languages spoken and linguistic proficiency. Cummins' *Interdependence Hypothesis* in bilingualism, for example, claims that second language (L2) skills are dependent in part on L1 language abilities, and that language skills will transfer from L1 to L2 given sufficient exposure and level of proficiency in L2. Thus, well-developed skills in L1 should support the development of similar skills in L2 (Cummins, 1979; Verhoeven, 1994). Within a broader framework of third language acquisition, it has been suggested that third language learners have a larger linguistic repertoire than second language learners, have stronger metalinguistic skills (Jessner, 2008; Huang, 2018), are more experienced with language acquisition processes, and thus, may have developed a unique set of language learning strategies that they can access (Jessner, 2008; Cenoz, 2013). Moreover, as a result of their wider linguistic repertoire, multilinguals may activate both direct routes and indirect routes to additional language learning (Hirosh and Degani, 2018). According to Hirosh and Degani (2018), direct routes include transfer of linguistic skills and knowledge, whereas indirect routes represent cognitive factors such as metalinguistic awareness and working memory.

In addition to learner attributes, cross linguistic relationships are also affected by properties of the languages in contact, such as general typological proximity or more specifically structural similarities. According to the *Linguistic Proximity Model* (Westergaard et al., 2017), structural similarities among the languages of a multilingual learner may lead to facilitative cross linguistic influences, particularly when the learner is aware of fine-grained variations across the grammars of their language repertoire. Non-facilitative influences may occur when the learner does not have a solid grasp of particular linguistic input in the target language and erroneously attributes shared properties between the target language and any of the already learned languages. Within this model, the specific patterns of influence will be determined by the areas of cross language

overlap (Westergaard et al., 2017). However, recent research indicates that there is a bidirectional transfer of skills from L1 to L2 (*forward transfer*) and from L2 to L1 (*reverse transfer*) (Jarvis and Pavlenko, 2008; Kim and Piper, 2019). Pavlenko and Jarvis (2002) examined bidirectional influences of L1 (Russian) and L2 (American English) oral production and found that not only did Russian influence English, but there was also a direct influence of English (L2) on L1 (Russian) oral production. Kim and Piper (2019) found bidirectional influences between Kiswahili and English (official languages in Kenya) literacy skills. The present study examines cross linguistic influences in EFL oral narrative production among monolingual Hebrew speakers and multilingual Arabic speakers.

## Oral Narratives

Oral narratives have long been recognized as a valid measure of linguistic growth in monolingual and bilingual children from different linguistic backgrounds (Berman and Slobin, 1994; Pavlenko, 2002; Soodla and Kikas, 2010). The interest in narratives is fueled by different disciplines: linguistics, sociology, cognitive psychology (Iluz-Cohen and Walters, 2012). Producing a narrative requires the integration of different cognitive and linguistic skills, e.g., interaction of lexical, morphosyntactic, and general discourse knowledge (Boerma et al., 2016), along with metacognitive skills (Cortazzi and Jin, 2007; Kupersmitt et al., 2014).

Growing narrative abilities also coincide with the development of executive functions, in this instance the ability to plan, i.e., organizing the story in sequential order (Friend and Bates, 2014), making narratives a perfect vehicle to access linguistic and cognitive growth in children. Expressing one's own and others' perspectives through lexical diversity and application of appropriate morphosyntactic structures (Moonsamy et al., 2009) involves both production and comprehension of multiple utterances and represents the current level of linguistic and conceptual knowledge of an individual (Justice et al., 2010). For these reasons, narratives have been examined from the point of view of (1) global characteristics, i.e., macrostructure, or producing a narrative based on the understanding of the thematic orientation, as it accounts for the "mental representation of events" (Berman, 1995, p. 287), and (2) overall grammatical complexity, e.g., microstructure, which includes lexical diversity as well as morphosyntactic knowledge, to represent the meaningful use of grammatical structures to allow listener's understanding (Justice et al., 2010). Moreover, strong correlations were found between these two structures of narratives, indicating that better lexical and morphosyntactic knowledge results in better global representation of narrative (Terry et al., 2013). Research also suggests that there is a strongly implied interaction between lexical and grammatical knowledge across different languages, such as Italian, Hebrew, Icelandic, etc. (Thordardottir et al., 2002). Moreover, Thordardottir et al. (2002) suggested that this interaction is due to a single mechanism which supports the development of both lexical and grammatical knowledge. Speakers of multiple languages also showed interaction between lexical and grammatical knowledge, when they produce narratives in their non-native languages

(Marchman et al., 2004; Simon-Cereijido and Gutiérrez-Clellen, 2009; Gagarina et al., 2015; Gagarina, 2016). However, there is no consensus regarding which particular factors, e.g., lexical, morphosyntactic, or general cognitive skills, may play mediating roles among learners of foreign languages in general, and among learners of EFL specifically, in their attempts to produce cohesive oral narratives in English. Despite the proposed interaction between lexical and morphological knowledge, there is also evidence that lexical and morphological knowledge may separate among second language learners, depending on the morphological tasks' reliance on general vocabulary (Shahar Yames et al., 2018).

Since the time when Labov (1972) delineated structural and functional aspects of narratives as macro- and micro-structure, or cohesion and coherence, researchers have consistently used specific markers to account for the completeness of narratives. The macrostructure assessment includes story setting (time and place), identifying the protagonist, references (use of pronouns with antecedents), and attempts to solve the problem (Pearson, 2002; Uccelli and Páez, 2007; Heilman et al., 2010). At the same time, when producing an oral text, a narrator in EFL must also show the ability to use appropriate lexical and morphosyntactic structures (microstructure, or coherence) in the non-native language in order for the narrative to be understood by the listener (Iluz-Cohen and Walters, 2012). This includes extensive vocabulary, appropriate knowledge of morphological inflections, as well as sentence complexity, e.g., use of cohesive devices, such as conjunctions. While research showed that the stories produced by multilingual speakers in their L1 and L2 may not really differ in the macrostructure, there were substantial differences in the appropriate use of the linguistic elements in L2 production (Pearson, 2002). Moreover, neither Pearson (2002), nor Uccelli and Páez (2007) later found any cross-linguistic transfer in children's narratives. On the other hand, Castilla et al. (2009) did find correlations across L1 and L2 (Spanish and English in this instance) in morphosyntactic knowledge, which indicated that there are some aspects of linguistic knowledge that actually can be transferred cross-linguistically. More recent research with Cantonese-English speaking preschoolers showed that Cantonese micro- and macrostructure predicted English micro- and macrostructure, again suggesting cross-linguistic transfer, as well as bidirectional influences (Rezzonico et al., 2016). This particular study is important, as it examined cross-linguistic transfer among typologically different languages. There is also substantial evidence that the transfer can occur on the level of morphosyntax, e.g., development of relative clauses in Cantonese-English bilinguals (Yip and Matthews, 2007), as well as in morphological awareness. For example, Pasquarella et al. (2011) showed bidirectional transfer of morphological awareness between English compound morphology and Chinese vocabulary acquisition, as well as English compound morphology predicting Chinese reading comprehension. However, the abovementioned studies examined written language. Nevertheless, based on these results, we may infer which L1 skills might contribute to producing L2 narratives among bilingual individuals, and individuals who are learning EFL with typologically different L1s. In other words, it is possible to postulate that there is a mediating

effect of specific L1 skill that will account for cross-linguistic transfer even in the oral language domain.

## Language Typology: Structural Similarities and Differences Between Arabic/Hebrew and English

Arabic and Hebrew are Semitic languages. As such, they are both morphologically rich with words being formed by combining roots comprised of two to four letters with patterns or templates: these are usually inserted between the root letters in order to create words with meaning. Some words are also formed by affixation at the beginning or end of the word. Verbs are inflected for number, gender, person (first, second, and third), and tense, or aspect to create three tenses: past, present, and future. Nouns are inflected for number and gender. Pronominal suffixes can be attached to verbs and nouns, thereby adding an additional layer of morphological complexity to words, so that most words are multi-morphemic. Research indicates that as a result of the morphological complexity of the language, Arabic and Hebrew speaking children show signs of morphological sensitivity at an early age and are well able to attend to internal word structure (Ravid, 2001; Gillis and Ravid, 2006; Saiegh-Haddad and Taha, 2017; El Akiki and Content, 2020). Moreover, there is empirical evidence that young bilingual Hebrew and Arabic speakers outperform monolinguals on tasks of derivational morphology, as a sign of positive cross linguistic influences (Asli-Badarneh and Leikin, 2019). These findings suggest that cross-linguistic influences can be traced among languages that belong to the same typological group, such as Hebrew and Arabic (Schwartz et al., 2016; Asli-Badarneh and Leikin, 2019). It has been further suggested that since Arabic morphology is more complex than Hebrew morphology, Arabic speaking bilinguals may have an advantage when they transfer skills from their more complex L1 morphological system to the less complex L2 Hebrew system (Chen and Schwartz, 2018). However, despite shared linguistic and structural features between Arabic and Hebrew, Arabic is unique in that it is a diglossic language, characterized by linguistic distance between the spoken dialects and the Modern Standard form (Saiegh-Haddad, 2003). Moreover, there may be great variance across spoken dialects in the areas of phonology, morphology, and semantics (Saiegh-Haddad and Henkin-Roitfarb, 2014), while the written form is stable. Children begin to learn the Modern Standard form formally when they enter school, although they are exposed to it informally in their environment from an early age through literature and media. Thus, Arabic speakers study academic subjects in school in Modern Standard Arabic (MSA), a language that is not their home language (spoken dialect). As a result of their exposure to two forms of the language from an early age, some researchers suggest that Arabic speaking children are, in essence, can be considered bilingual (Eviatar and Ibrahim, 2000), e.g., the spoken form is their L1 and the Modern Standard form is their L2, even before they begin to study Hebrew. Studies have also indicated that the linguistic distance between the two forms of Arabic language may result in difficulties with the acquisition of the Modern Standard form,

and other literacy related skills (Saiegh-Haddad, 2003, 2007). Moreover, these difficulties persist beyond the initial years of literacy acquisition (Abdelhadi et al., 2011).

In contrast to Semitic languages, English is a Western Germanic language with a relatively simple morphological structure. Morphemes are affixed onto base words producing multimorphemic words, which can be inflectional or derivational in nature. Inflectional morphemes change tense or number of free morphemes, whereas derivational morphology can change the part of speech or the meaning of the base word. In addition to the three basic tenses (past, present, future), English verb tenses are complicated by the use of aspectual forms, which contribute an additional level of linguistic complexity. And yet, English has been described as a morphologically impoverished language due to the fact that, in comparison with Semitic languages, there are hardly any markings of agreement for gender, number, or person (DeKeyser et al., 2010; Tsarfaty and Sima'an, 2010).

In addition, word order within a sentence in Semitic languages, is rather flexible. This is because syntactic characteristics are determined based on morphological information embedded within words, above and beyond word order within a sentence (Tsarfaty and Sima'an, 2010). Thus, while the dominant word order for Hebrew sentences is SVO, and VSO for Arabic sentences, variability in word order appears and is accepted in both languages (Schwartzwald, 2011). Moreover, verbless constructions exist in both languages. In contrast, English word order within sentences is relatively fixed according to the SVO pattern so that grammatical patterns are reliant on syntax along with meaning. The present study examines oral language skills in EFL among Arabic and Hebrew speaking pupils in sixth grade, thus, these particular differences between Hebrew, Arabic and English language structure provide an important frame of reference.

## Importance of English and EFL Learning Policy in Israel

English is a global language (Crystal, 2012). It is also the most widely taught foreign language today. In Israel, English holds a unique status. It is a semi-official language and the first foreign language studied in all schools. It is the key to international economic growth, and social communication, as well as the gatekeeper to academic advancement. It is also a compulsory subject for the high school matriculation certificate. Moreover, English proficiency is required as a prerequisite for entrance into higher education.

Based on the language learning policy of the country, Arabic speakers are taught in school in Arabic, and Hebrew speakers are taught in Hebrew. Arabic speaking pupils must also learn Hebrew as their second written and third spoken language, beginning in the second grade or even earlier (Amara, 2018). Hebrew speaking pupils only begin to study Arabic in the 7<sup>th</sup> grade, if at all. Formal instruction in English as the first foreign language (EFL) for both populations begins in 3<sup>rd</sup> grade.<sup>1</sup> Thus, while Hebrew speakers study EFL as their second language,

<sup>1</sup>[https://meyda.education.gov.il/files/Mazkirut\\_Pedagogit/English/curriculum2020Elementary.pdf](https://meyda.education.gov.il/files/Mazkirut_Pedagogit/English/curriculum2020Elementary.pdf)

native Arabic speakers study EFL as the 3<sup>rd</sup> written language and fourth spoken one. Moreover, the status of each of the languages is evident from their appearance in the linguistic landscapes of the community and the school settings. Among the Hebrew speaking communities, Hebrew is dominant in the linguistic landscape, followed by English. In most Arabic speaking communities, Arabic is the dominant language in the linguistic landscape followed by Hebrew. English is hardly present, except for the region of East Jerusalem where it is the second most prominent language after Arabic (Amara, 2018). A similar situation exists within the school linguistic landscapes, at least in the Arabic speaking population, where Arabic is the most prominent language followed by Hebrew with little or no representation of English (Amara, 2018).

Based on the national guidelines, all pupils in Israel learn EFL according to the same English curriculum and teaching materials, regardless of their L1 backgrounds. The first year of study (3<sup>rd</sup> grade) is dedicated to building basic oral skills and introducing the letters of the alphabet. Pupils in 3<sup>rd</sup> grade study English for two 45-min lessons a week. From 4<sup>th</sup>–6<sup>th</sup> grades pupils receive four 45-min EFL lessons a week. In the 4<sup>th</sup> grade, pupils continue to build their lexical knowledge while beginning to learn how to read. By the last year of elementary school (6<sup>th</sup> grade), pupils are expected to reach a basic level of oral and written language proficiency.

In reality, however, after the first year of oral language instruction, there is little, if any time specifically devoted to the further explicit development of authentic oral language skills (Al Hosni, 2014). While the revised elementary English curriculum<sup>1</sup> relates to oral language skills, the expectation is that the pupils will be able to manage short, rehearsed utterances, and not necessarily produce spontaneous language. Moreover, the reality of large numbers of pupils in each class does not really allow for individual practice or use of oral language skills in school. Additionally, while the revised curriculum highlights the importance of English oral skills in the classroom, many teachers prefer to use their L1 (Orland-Barak and Yinon, 2005; Timor, 2012), which further limits exposure to spoken English.

## Present Study

The present study explores the contribution of cross linguistic influence to oral narrative production in EFL among a sample of Hebrew and Arabic speakers. To our knowledge, this is the first study that directly examines cross language influences and possible specific influential factors between Semitic L1 and EFL through oral narratives. Previous studies concerned with narrative production among bilingual individuals compared macro- and micro-structures in relation to the differences of the output between L1 and L2 of participants (Uccelli and Páez, 2007; Gagarina, 2016; Lucero, 2018). However, there was no specific data relating to which L1 skill(s) had a mediating effect on L2 narratives, as well as the possibility that different EFL skills may prove to be more influential for oral EFL production among Arabic and Hebrew speakers. The present study addresses this gap by examining oral narratives produced by 6<sup>th</sup> grade speakers of Arabic and Hebrew in their 4<sup>th</sup> year of learning EFL, within a cross-linguistic framework, in search for possible

mediators between Semitic languages and English. For these purposes, we did not compare L1 and EFL narratives, but rather examined the proficiency of the participants in linguistic and literacy domains in L1 Hebrew and L2 Modern Standard form of Arabic (MSA) and EFL, as well as general cognitive skills, as possible contributing factors to the quality of EFL oral narratives.

Our research questions were as follows:

1. What is the contribution of L1 Hebrew linguistic skills (morphology, reading comprehension), as well as phonological memory (PM) (word repetition), and general cognitive abilities, to micro- and macro-structures in EFL oral narratives among monolingual Hebrew speakers?

We hypothesized that in line with the *Interdependence hypothesis* (Cummins, 1979, 2014) we should see direct cross-linguistic influence with strong L1 skills supporting similar skills in L2. However, if non-facilitative influences are observed, these findings may be attributed to either typological differences between Hebrew, a Semitic language, and EFL, as postulated by the *Linguistic Proximity model*, or low levels of EFL proficiency.

2. What is the contribution of MSA and Hebrew (L3) linguistic skills (morphology, reading comprehension), as well as cognitive skills (phonological and general cognitive abilities), to micro- and macro-structures in EFL oral narratives among multilingual Arabic speakers?

We hypothesized that since the Arabic speakers in the present study are multilingual and have a broader linguistic repertoire than the monolingual Hebrew speakers, there may be a different allocation of cognitive and linguistic resources that will impact the nature of the cross linguistic influences on EFL narrative skills. Moreover, there may be both direct transfer of linguistic skills and knowledge between the same Arabic and Hebrew skills, as well as indirect transfer of cognitive and metalinguistic skills to EFL narratives, as long as their Arabic and Hebrew linguistic skills are at a sufficient level of proficiency to support transfer.

3. What is the contribution of EFL linguistic skills [English reading comprehension (ERC), English morphological awareness (EMA)] to micro- and macro-structures in EFL oral narratives among monolingual Hebrew speakers and multilingual Arabic speakers?

We hypothesized that as macrostructure is a measure of the overall ability to represent global characteristics of a narrative, ERC scores, representing an understanding of the story events, would have a positive relationship with EFL macro- and Total narrative scores. We further hypothesized that since microstructure is a measure of lexical diversity as well as morphosyntactic knowledge (Justice et al., 2010), EMA scores would have a positive relationship with EFL micro- and Total narrative scores. In the event that the abovementioned EFL linguistic skills show a positive relationship with all EFL narrative measures, it will be possible to infer that these skills may play a modulating role in the relationship between Arabic, Hebrew, and EFL narratives.

4. Are there L1(Hebrew)/L2 (MSA)/L3 (Hebrew for Arabic speakers) skills that may have cross linguistic influences in the relationship between EFL skills and EFL oral narratives?

**TABLE 1** | Descriptive statistics in percentage scores (except for Arabic morphological word derivation task which are presented as raw scores).

| Task                                                          | Hebrew speakers     | Arabic speakers     |
|---------------------------------------------------------------|---------------------|---------------------|
|                                                               | Mean (SD)<br>N = 86 | Mean (SD)<br>N = 85 |
| Microstructure                                                | 17.64 (15.18)       | 12.16 (12.79)       |
| Macrostructure                                                | 47.83 (20.15)       | 36.86 (19.40)       |
| Total narrative                                               | 26.47 (15.56)       | 19.38 (13.79)       |
| Raven                                                         | 87.58 (11.5)        | 79.89 (13.06)       |
| English word repetition (as a measure of phonological memory) | 91.26 (9.83)        | 93.08 (8.24)        |
| Arabic morphological word derivation <sup>a</sup>             | X                   | 15.32 (5.14)        |
| Arabic morphological root pattern awareness                   | X                   | 89.23 (15.69)       |
| Arabic reading comprehension                                  | X                   | 76.34 (23.31)       |
| Hebrew morphological real word derivation                     | 83.59 (14.23)       | 16.93 (12.96)       |
| Hebrew morphological pseudo word derivation                   | 77.23 (14.77)       | 26.61 (22.07)       |
| Hebrew reading comprehension                                  | 80.50 (17.19)       | X                   |
| English morphological awareness                               | 65.36 (21.44)       | 44.97 (29.77)       |
| English reading comprehension                                 | 61.01 (21.44)       | 43.37 (26.19)       |

<sup>a</sup>This score is reported as a raw score due to the nature of the task.

We hypothesized that possible cross linguistic relationships between measures of EFL (ERC and EMA) and L1 Hebrew/L2 MSA/L3 (Hebrew for Arabic speakers) MA and RC, as well as EFL total narrative structure should indicate an indirect mediation effect on the quality of the EFL narrative.

## METHODS

### Participants

The participant pool was comprised of two linguistic groups: 86 native mono-literate sixth grade Hebrew speaking pupils for whom English is the second language being learned in school (43 females), and 85 native 6<sup>th</sup> grade Arabic speaking pupils, who learn spoken language as L1, MSA as L2, learn Hebrew as L3 and then begin learning EFL. Therefore, this group can be considered “multilingual,” as English is third written language and the fourth spoken language being learned in school (44 females). All of the participants were between the ages of 10 and 11 at the time of data collection. The Arabic speaking pupils were chosen from four different schools, and the Hebrew speaking pupils were chosen from six different schools in the central area of Israel. As the Arabic speaking participants were from several different cities, their spoken dialects varied in relation to the city where the participants lived. The average socio-economic index for the Arabic speaking school was 5.66, and for the Hebrew speaking schools 2.95 on a scale of 1–10 where 1 is the highest and 10 is the lowest. However, very large discrepancies can be found within the Arabic speaking populations that were included in the study. As a case in point, in our groups, there were significant individual differences on task results, as can be seen in the very large standard deviations in **Table 1**<sup>2</sup>. The protocol of this study was

approved by the Ministry of Education Chief Scientist Bureau. Pupils chose to participate on a voluntary basis and all parents signed a consent form. Pupils with learning disabilities or pupils who are fluent in languages in addition to Hebrew or Arabic were excluded from the sample.

### Tasks

Language tasks in Arabic and Hebrew were adapted from existing standardized tools (Arabic—Asadi et al., 2015; Hebrew—Shany et al., 2006). All the Arabic tasks were administered in the MSA (which, as mentioned earlier, is considered the first written language that the Arabic speakers acquire but the second spoken language). Arabic speakers were also tested in Hebrew—their second written and third spoken language. While most of the Arabic and Hebrew tests have norms, some of the tasks were shortened for the purposes of the present study and therefore the norms can only be used to give an indication of performance trends. Further, the Hebrew tasks were not designed to be administered to non-native Hebrew-speaking pupils, so the scores for the Arabic speakers cannot be evaluated with relation to the Hebrew norms. All tests, in both languages, were administered by native speakers of the language of the participants. In what follows, the tasks will be described, according to language of administration.

### Raven Matrices (As a Measure of General Cognitive Ability)

Raven Colored Matrices (Raven et al., 1976), as a test of non-verbal intelligence, is comprised of three sets of tasks with 12 items in each set (maximum possible score = 36). The participant is required to select the correct pattern to complete a matrix out of six options. Prior to testing, participants received an explanation as to how to fill out the answer sheet. This task was administered in Arabic for the Arabic speakers and Hebrew for the Hebrew speakers (Cronbach's Alpha 0.87).

<sup>2</sup>In this study, we are comparing the mean performance of each group and not individuals. Since performance by each group on the native language tasks was at least within the high average range on all native language tasks, we do not consider the differences in SES to be an obstacle.

## Arabic Assessment Battery

### *Reading Comprehension (Asadi et al., 2015)*

This task requires the participants to read a short passage (205 words in length) and answer nine multiple choice questions. The expository text is about a subject that is relevant to children in the sixth grade (Cronbach's alpha 0.71).

### *Morphological Choice (Asadi et al., 2015)*

This task examines students' awareness of the roots in words using an *odd one out* format. The pupil is presented with four written words in a row and is required to choose the item that does not belong based on a change to the root. Roots are changed in the words by switching the order of the first or second letter. There are a total of 20 sets of four items (Cronbach's alpha 0.90).

### *Morphological Word Derivation (From Roots) (Asadi et al., 2015)*

In this task, the pupil is presented with five three-letter roots and has 1 min to derive as many words from each root as possible. One point is given for each correctly derived word (Cronbach's alpha 0.82).

Arabic speakers were also assessed on Hebrew MA (real and pseudo word derivations), as they learn Hebrew as their second written language (see descriptions of the tasks below).

## Hebrew Assessment Battery

### *Reading Comprehension*

Participants were presented with a reading passage (218 words) which was adapted from a sixth grade Hebrew textbook. The expository text is about a subject that is relevant to children in the sixth grade. There were nine multiple choice questions. Prior to administration, four Hebrew language teachers rated the text as grade appropriate (Cronbach's alpha 0.33).

### *Morphological Real Word Derivation (Shany et al., 2006)*

This task comprised nine sentences where one word was missing from each sentence (*cloze* procedure). The pupils were given a three-letter root pattern and were required to fill in the missing word in the sentence by deriving the correct form of the word in the context of the given sentence (Cronbach's alpha 0.91).

### *Morphological Pseudo Word Derivation (Shany et al., 2006)*

This task comprised 12 sentences where one word was missing from each sentence (*cloze* procedure). The pupils were given a three-letter pseudo root pattern and were required to fill in the missing pseudo word in the sentence by deriving the correct form of a word in the context of the given sentence. Success on this task was dependent on an understanding of the morphological patterns governing word construction in Hebrew. There were 12 items in this task (Cronbach's alpha 0.89).

## English Assessment Battery

### *Word Repetition (As a Measure of Phonological Memory)*

Pupils heard a word and were asked to repeat the word after the tester. Twenty-five items for this task were chosen based on number of syllables (2–4) and level of familiarity (None of the chosen words appear in the first 1,200 words from the list of lexical items of the English Inspectorate of the Ministry of

Education, Israel. Thus, as the chosen words were unfamiliar to the participants, they were, in essence, pseudo words). This task tapped into phonological short-term memory (Cronbach's alpha 0.66).

### *Reading Comprehension*

This task included a passage (101 words), composed of vocabulary taken from the list of the first 1,200 words proposed by the English Inspectorate of the Ministry of Education, Israel (State of Israel, Ministry of Education Pedagogical Secretariat, Language Department, Inspectorate for English Language Education, 2020), followed by seven multiple choice questions. The topic was chosen by reviewing the content of the textbooks used in the fifth and sixth grades. Once the text and questions were written, the passage was given to three different elementary school English teachers who were asked to judge if the passage and questions were appropriate for the sixth graders. Each of the teachers assessed the passage as suitable. This task also served as a proxy for EFL vocabulary task, because in order to understand a written text, one must have a substantial knowledge and comprehension of vocabulary items (Cromley and Azevedo, 2007) (Cronbach's alpha 0.73).

### *Morphological Awareness*

This task comprised nine sentences where one word was missing from each sentence (*cloze* procedure). The pupils were given a base word and were required to fill in the missing word in the sentence by deriving the correct form of the base word in the context of the given sentence. Morphological structures targeted included both inflectional and derivational morphemes (for example: I teach English to my pupils. I am a \_\_\_\_.) (Cronbach's alpha 0.77).

## Oral Narrative Task in English (Dependent Variable)

All participants were asked to produce an elicited oral narrative in EFL based on the "Cookie Theft" task (Goodglass and Kaplan, 1983). This particular task presents extensive opportunity for children to describe the salient elements (e.g., mother washing the dishes, water overflowing from the tap, a boy, attempting to get cookies from a jar stored in a cupboard, the chair that is falling down, the girl stretching her hand to reach the cookies) as well as background features (e.g., trees, grass, clothes, etc.) of a black and white picture. All children had the prompt "Tell me about this picture" to orient them to all aspects of the picture, however, it is up to the narrator to relate to specific details. Despite its static nature, the picture presents the opportunity to use a variety of lexical items, including abstract and concrete words, and verbs in appropriate tenses, to construct grammatically correct sentences, and provide cohesive references (Cummings, 2019). Moreover, evidence suggests that narratives produced based on static picture show higher mastery than the ones produced on a series of pictures (e.g., famous "Frog Where Are You" by Mercer Myer) (Cornaglia et al., 2017).

The narratives were recorded and then transcribed by a trained research assistant and rechecked by the second author. The data was then analyzed using the items included in Narrative Assessment Protocol (short form, Justice et al., 2010) to assess

the microstructure of the narratives. The following items from the NAP short form were included to assess microstructure: number of complete sentences, number of complex sentences (e.g., sentences with clauses and conjunctions), instances of noun plural inflection, verb morphology, such as present progressive (e.g., auxiliary + main verb), copula *be*, 3<sup>rd</sup> person singular, past tense. The macrostructure analysis, i.e., completeness of the narratives, was adapted from Narrative scoring scheme (Heilman et al., 2010). In what follows the method of assessment will be described, first for micro- and then for macro-structure.

The microstructure analysis was comprised of two indices representing cohesion: (1) The “*Complete Sentences*” index which was a combination of all the required linguistic elements in their correct forms (i.e., use of appropriate inflectional verb morphology and maintenance of the English word order); and (2) the “*Conjunction Cohesion*” index, which was a measure of sentence complexity, e.g., use of clausal structures. Each instance of full sentence, clausal structure and correct use of inflectional morphology received a score of 1. The maximum total possible score (sum of “*Complete Sentences*” and “*Conjunction Cohesion*”) for the microstructure was 36, similar to the one suggested by the NAP short form, where total instances of each of the correctly produced required elements could not be assigned a score higher than 3 (the short form uses 3+ for the highest possible frequency of use) (Cronbach  $\alpha = 0.57$ ). This arrangement was deemed appropriate for the purposes of this study, as many students produced only labels for the items they saw in the picture. However, many nouns used by participants were inflected for plurality, allowing the student to receive a score.

To assess the macrostructure, we modified the narrative scoring scheme suggested by Heilman et al. (2010). Since we used static stimulus, the picture was divided into three episodes: (1) Mother (washing dishes, holding plate, drying dishes, looking out of the window, etc.); (2) boy/girl (reaching for cookies, giving cookies, eating cookies, climbing chair, etc.); and (3) water overflowing from the sink (water on the floor, wet floor, mother/children don’t see it, etc.) with five indices each (e.g., Topic maintenance, Event Sequencing, Information, Referencing, Character ID) scored on a 0–3 scale, where 0 signified non-observed ability, resulting in a total of 15 points (Cronbach  $\alpha = 0.74$ ). The rater agreement (Cohen’s kappa, Cohen, 1960) for the transcription and scoring was at 0.87, which is “almost perfect” (82% and above is reliable) (see **Appendix 1** for a detailed description).

## Data Analysis

The present study set out to explore the contribution of cognitive and linguistic skills to the production of oral narratives in EFL, among L1 Hebrew and Arabic speakers in 6<sup>th</sup> grade, as well as identify the underlying mechanisms that lead to cross-linguistic transfer and the mediating factor(s) that may be associated with this process. We included descriptive statistics of all test results in both linguistic groups, which are represented through means and standard deviations. Interpretations of the scores were done in accordance with norms for tasks that were norm-referenced. The main part of the analysis was based on hierarchical multi-variate linear regression models. Prior to the named analyses

assumption of linearity relationship was tested by curve fitting analysis that tested the significance of the linear relationships between the predictors in the models and the outcome variable within each language group. This analysis revealed that these bivariate associations had significant linear relationship with the outcome variable.

The hierarchical modeling strategy was based on the following logic: first we monitored general cognitive abilities, represented in the study by Raven Colored Matrices and by English word repetition (EWR) (cognitive/linguistic task for assessing PM). Then, we wanted to see the additional contribution of language tasks in first and additional languages. For Hebrew speakers, this meant all tasks in Hebrew first and then in EFL. For Arabic speakers, Arabic language measures (MSA) were entered first, followed by language measures in Hebrew (L3) and then in EFL (L4). In terms of the final structure of multi-variate linear regression, there were three models for Hebrew speakers and four models for Arabic speakers. The Durbin-Watson test for independence of errors, showed values for both language groups were between 1.50 and 2.00, which indicates that residuals are uncorrelated.

The investigation of the L1/L2/L3 skills that support EFL oral narratives used the assumptions of the *Confounding hypothesis* (MacKinnon et al., 2000), that a 3<sup>rd</sup> variable (Z) may explain the relationship between the predictor (X) and outcome variable (Y), by having an impact on both. We include the results of the bootstrapping and Sobel Test procedures, as the statistical analysis for mediation and confounding are similar, except for assumption of the directionality (Hayes, 2019). All analyses were performed using SPSS25, results were considered significant when  $p \leq 0.05$ .

## RESULTS

The present study set out to explore the contribution of cognitive and linguistic skills to the production of oral narratives in EFL, among native Hebrew and Arabic speakers in 6<sup>th</sup> grade, as well as identify the underlying mechanisms that lead to cross-linguistic transfer. We first established base-line Arabic, Hebrew and EFL proficiency scores in order to validate the assumptions of the *Interdependence Hypothesis*, that suggests reliance of additional language acquisition on native language proficiency.

**Table 1** represents the percentage scores on all tasks, except for the score on the Arabic morphological derivation, which was reported as a raw score because of the nature of the task, where there was no definite number for the total possible score. The Arabic speakers were tested in both Arabic (MSA) and Hebrew (L3), before being tested in EFL. **Appendix 2** presents a list of all the abbreviations for tasks reported in the results and discussion sections.

As can be seen in **Table 1**, each group showed a unique psycholinguistic profile. While the Arabic reading comprehension (ARC) and morphological choice tasks were shortened versions of the same tasks from the standardized Arabic test battery (Asadi et al., 2015), the scores in **Table 1** all fall within the normative range as compared to the scores for the similar full tasks. The scores on the adapted version of the

**TABLE 2 |** Hebrew speakers: correlations among all measures and English oral narrative scores.

| Variables                | 1      | 2      | 3      | 4    | 5      | 6      | 7      | 8    | 9      | 10 |
|--------------------------|--------|--------|--------|------|--------|--------|--------|------|--------|----|
| 1. E Micro               | –      |        |        |      |        |        |        |      |        |    |
| 2. E Macro               | 0.73** | –      |        |      |        |        |        |      |        |    |
| 3. Total narrative       | 0.97** | 0.88** | –      |      |        |        |        |      |        |    |
| 4. Raven                 | 0.01   | 0.04   | 0.02   | –    |        |        |        |      |        |    |
| 5. HMA <sub>real</sub>   | 0.22*  | 0.22*  | 0.23*  | 0.04 | –      |        |        |      |        |    |
| 6. HMA <sub>pseudo</sub> | 0.25*  | 0.26*  | 0.27*  | 0.19 | 0.40** | –      |        |      |        |    |
| 7. H RC                  | 0.20   | 0.15   | 0.19   | 0.13 | 0.05   | 0.18   | –      |      |        |    |
| 8. E WR                  | 0.27*  | 0.35** | 0.32** | 0.16 | 0.26*  | 0.26*  | 0.27*  | –    |        |    |
| 9. EMA                   | 0.26*  | 0.26*  | 0.28*  | 0.03 | –0.06  | 0.07   | 0.03   | 0.07 | –      |    |
| 10. ERC                  | 0.48** | 0.52** | 0.53** | 0.04 | 0.13   | 0.35** | 0.31** | 0.18 | 0.47** | –  |

HMA<sub>real</sub>, Hebrew morphological real word derivation task; HMA<sub>pseudo</sub>, Hebrew morphological pseudo word derivation task; RC, reading comprehension; H, Hebrew; E, English; MA, morphological awareness; WR, word repetition.

\* $p < 0.05$ ; \*\* $p < 0.01$ .

morphological derivation task are also reflective of the normative scores from the standardized Arabic test battery (Asadi et al., 2015). Thus, it is possible to say that the Arabic speakers in the present study exhibited adequate normative proficiency in MSA. With regards to the performance of the Arabic speakers on the Hebrew (L3) tasks, scores among the Arabic speakers were higher on the Hebrew morphological pseudo word derivation task than on the morphological real word derivation task. Scores on the EFL tasks fell within the moderate range of proficiency.

The scores for the Hebrew speakers on the two Hebrew (L1) morphological awareness tasks were within the average high to high range, based on the given norms for those tasks (Shany et al., 2006). While the reading comprehension task was not norm-referenced, a mean score of 80 could be considered within the high average range. Scores for the EFL tasks among the group of Hebrew speakers fell in the average high range of proficiency.

## Research Questions 1, 2, and 3

The first research question addressed the contribution of different cognitive and linguistic skills in Hebrew, as L1 to the ensuing EFL narratives. Our second research question addressed the contributions of cognitive and linguistic skills in Arabic as the first written language of the participants, along with Hebrew as the second written language and third spoken language acquired by Arabic speakers, and micro- and macro-structures of EFL narratives. The third question explored the contribution of EFL linguistic skills (ERC, EMA) to EFL oral narratives among the monolingual Hebrew and multilingual Arabic speakers. In what follows, results will be presented for each language group separately.

## Hebrew-Speaking Participants

### Correlation Analysis

In this section, we present only the most significant correlations to identify the variables that were used in *Regression Analysis*. *Raven*, as a measure of general cognitive abilities did not show correlations with any measure of Hebrew, EFL, and narrative components. However, EWR as a measure of PM, strongly and significantly correlated with every measure of

Hebrew assessments (two tasks of HMA and HRC), as well as every component of EFL narratives (see **Table 2** for results). Both, HMA<sub>pseudo</sub> and HMA<sub>real</sub> tasks strongly and significantly correlated with all elements of EFL narratives, with HMA<sub>pseudo</sub> being the strongest, and ERC. Moreover, HRC showed significant correlation with ERC. EFL tasks, e.g., EMA and ERC showed strong intra language correlations with each other and with all elements of narrative structure (see **Table 2** for all results).

### Hierarchical Regression

The total narrative score was used as the dependent measure, since it represents both, the micro- and macro-structures of the narrative. The independent variables were entered in an hierarchical fashion: the first block contained the Raven and EWR scores as representations of general cognitive skills (Model 1). In the second block the Hebrew language variables (both MA tasks and RC) were added (Model 2). In the third block the English language variables were added (Model 3). The final model was significant and explained 35% of the variance in the total narrative score [ $F_{(2,75)} = 10.10$ ,  $p < 0.001$ ]. Two predictors were found to have a significant association with the dependent variable: ERC ( $\beta = 0.40$ ,  $p < 0.001$ ) and EWR ( $\beta = 0.24$ ,  $p < 0.01$ ). No significant associations were found for the other variables in the model. The regression model for Hebrew is presented in **Table 3**.

## Arabic-Speaking Participants

### Correlation Analysis

Notable differences were observed in the relationships between the same cognitive and linguistic variables among Arabic-speaking children. As Arabic speakers learn Hebrew as their second written and third spoken language, Hebrew measures were included in the analysis, specifically to identify possible relations between Arabic and Hebrew morphological knowledge as part of the cross linguistic language profile. Firstly, *Raven*, as a measure of general cognitive skills, did not correlate with any of Arabic morphological awareness tasks (AMA<sub>choice</sub> and AMA<sub>deriv</sub>), but significantly correlated with ARC, HMA<sub>real</sub>, and all aspects of EFL narratives (**Table 4**). EWR, as a measure of

**TABLE 3 |** Hierarchal linear regression analysis for contributors to total narrative score among Hebrew speakers.

| Variable                                   | Model 1  |             |         | Model 2                           |             |         | Model 3                               |             |         |
|--------------------------------------------|----------|-------------|---------|-----------------------------------|-------------|---------|---------------------------------------|-------------|---------|
|                                            | <i>B</i> | <i>SE B</i> | $\beta$ | <i>B</i>                          | <i>SE B</i> | $\beta$ | <i>B</i>                              | <i>SE B</i> | $\beta$ |
| Raven                                      | −0.11    | 0.39        | −0.03   | −0.23                             | 0.40        | −0.06   | −0.15                                 | 0.36        | −0.04   |
| English word repetition                    | 0.56     | 0.17        | 0.35*** | 0.43                              | 0.18        | 0.27**  | 0.38                                  | 0.16        | 0.24**  |
| Hebrew morphological awareness pseudowords |          |             |         | 0.12                              | 0.13        | 0.11    | −0.01                                 | 0.12        | −0.01   |
| Hebrew morphological awareness real words  |          |             |         | 0.14                              | 0.12        | 0.13    | 0.15                                  | 0.11        | 0.14    |
| Hebrew reading comprehension               |          |             |         | 0.12                              | 0.11        | 0.12    | 0.03                                  | 0.10        | 0.04    |
| E morphological awareness                  |          |             |         |                                   |             |         | 0.07                                  | 0.08        | 0.09    |
| E reading comprehension                    |          |             |         |                                   |             |         | 0.21                                  | 0.06        | 0.40*** |
| $R^2$                                      |          | 0.12        |         |                                   | 0.17        |         |                                       | 0.35        |         |
| $R^2$ change                               |          | 0.12        |         |                                   | 0.05        |         |                                       | 0.18        |         |
| <i>F</i> for change in $R^2$               |          |             |         | $F_{(3,77)} = 1.58$<br>$p = 0.20$ |             |         | $F_{(2,75)} = 10.10$ ,<br>$p = 0.000$ |             |         |

\*\**p* is significant at 0.01; \*\*\**p* is significant at 0.001.

**TABLE 4 |** Arabic speakers: correlations among all measures and English oral narrative scores.

| Variables                | 1      | 2      | 3      | 4      | 5     | 6      | 7      | 8      | 9      | 10     | 11     | 12 |
|--------------------------|--------|--------|--------|--------|-------|--------|--------|--------|--------|--------|--------|----|
| 1. E Micro               | –      |        |        |        |       |        |        |        |        |        |        |    |
| 2. E Macro               | 0.70** | –      |        |        |       |        |        |        |        |        |        |    |
| 3. Total Nar             | 0.95** | 0.88** | –      |        |       |        |        |        |        |        |        |    |
| 4. Raven                 | 0.43** | 0.36** | 0.44** | –      |       |        |        |        |        |        |        |    |
| 5. AMA <sub>deriv</sub>  | 0.28*  | 0.25*  | 0.29** | 0.14   | –     |        |        |        |        |        |        |    |
| 6. AMA <sub>choi</sub>   | 0.20   | 0.23   | 0.23   | 0.05   | 0.25* | –      |        |        |        |        |        |    |
| 7. A RC                  | 0.35** | 0.33** | 0.37** | 0.40*  | 0.25* | 0.41** | –      |        |        |        |        |    |
| 8. HMA <sub>real</sub>   | 0.26*  | 0.28** | 0.30** | 0.22*  | 0.09  | 0.22   | 0.40** | –      |        |        |        |    |
| 9. HMA <sub>pseudo</sub> | 0.31** | 0.34** | 0.36** | 0.10   | 0.22* | 0.29** | 0.42** | 0.46** | –      |        |        |    |
| 10. E WR                 | 0.24*  | 0.31** | 0.29** | 0.201  | 0.09  | 0.39** | 0.32** | 0.21   | 0.23*  | –      |        |    |
| 11. EMA                  | 0.49** | 0.52** | 0.55** | 0.26*  | 0.24* | 0.29** | 0.09   | 0.21   | 0.44** | 0.38** | –      |    |
| 12. E RC                 | 0.46** | 0.38** | 0.46** | 0.34** | 0.26* | 0.16   | 0.42** | 0.33** | 0.37** | 0.24*  | 0.63** | –  |

AMA<sub>choice</sub>, Arabic morphological root pattern awareness task; AMA<sub>deriv</sub>, Arabic morphological word derivation task; HMA<sub>real</sub>, Hebrew morphological real word derivation task; HMA<sub>pseudo</sub>, Hebrew morphological pseudo word derivation task; RC, reading comprehension; H, Hebrew; E, English; MA, morphological awareness; WR, word repetition.

\**p* < 0.05. \*\**p* < 0.01.

PM, significantly correlated with AMA<sub>choice</sub> and all aspects of narrative structure. AMA<sub>choice</sub> also significantly correlated with ARC, EMA, macro- and total narrative structure, while AMA<sub>deriv</sub> correlated with all EFL skills, e.g., EMA and ERC, as well as with all aspects of narrative structures. Moreover, we saw significant correlations between HMA<sub>pseudo</sub> and both AMA tasks. Similar to the Hebrew sample, ARC showed strong correlation with ERC. As with the Hebrew speakers, we saw significant inter-correlations between micro- and macro-structures as well as with Total score among Arabic-speaking children (see Table 4 for all results).

### Regression Analyses

The Total Narrative score was again our dependent measure, however, in addition to MSA measures, Hebrew L3 tasks were added to the models. The predictor models for EFL narratives among Arabic speakers yielded distinctly different results from the Hebrew models. Model 1 included the same predictors as for Hebrew speakers, e.g., cognitive variables. Arabic linguistic tasks

(MA and RC) were added to Model 2. In Model 3 we added the two Hebrew measures of MA, as L3 skills for Arabic speakers. EFL variables of ERC and EMA were entered as Model 4 (see Table 5). The final model was significant and explained 42% of the variance in the total narrative score [ $F_{(2,72)} = 5.25$ ,  $p < 0.001$ ]. Two predictors were found to have a significant association with the dependent variable: Raven ( $\beta = 0.29$ ,  $p < 0.01$ ) and EMA ( $\beta = 0.33$ ,  $p < 0.01$ ). No significant associations were found for the other variables in the model. The regression model for Arabic is presented in Table 5.

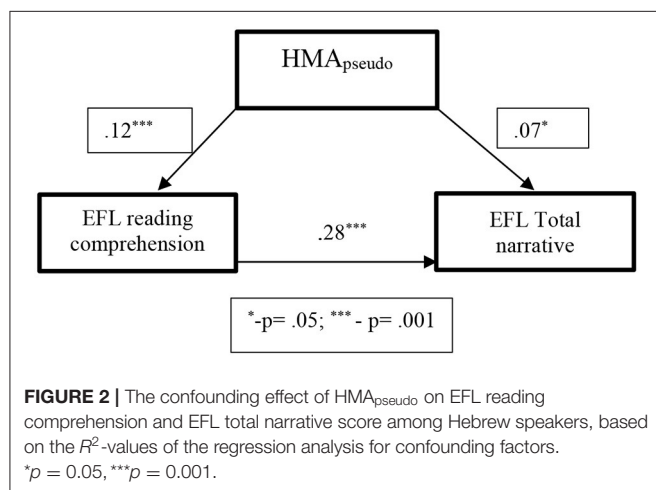
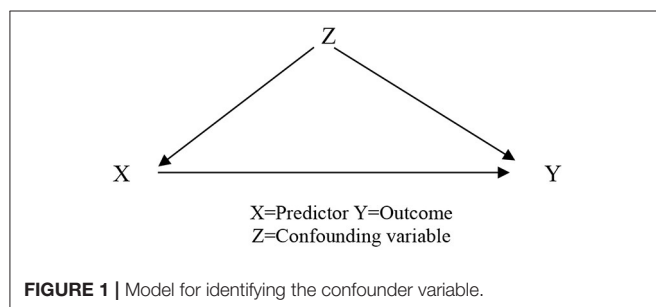
### Research Question Four

Our final research question addressed the possible role of L1/L2/L3 skills in the production of EFL narratives. In essence, we were looking for a specific L1/L2/L3 skill that may give additional support to the quality of EFL oral narratives, based on the MacKinnon et al. (2000) suggestion that after establishing the relationship between predictor and outcome variables it is common to identify a third variable (e.g., specific skill) that may

**TABLE 5 |** Hierarchal linear regression analysis for contributors to total narrative score among Arabic speakers.

| Variable                                    | Model 1 |      |         | Model 2 |                                   |         | Model 3 |                                   |         | Model 4 |                                   |         |
|---------------------------------------------|---------|------|---------|---------|-----------------------------------|---------|---------|-----------------------------------|---------|---------|-----------------------------------|---------|
|                                             | B       | SE B | $\beta$ | B       | SE B                              | $\beta$ | B       | SE B                              | $\beta$ | B       | SE B                              | $\beta$ |
| Raven                                       | 1.20    | 0.32 | 0.38*** | 1.04    | 0.32                              | 0.33*** | 1.04    | 0.31                              | 0.33*** | 0.92    | 0.31                              | 0.29*** |
| English word repetition                     | 0.37    | 0.17 | 0.22**  | 0.25    | 0.18                              | 0.14    | 0.21    | 0.18                              | 0.13    | 0.07    | 0.17                              | 0.04    |
| Arabic morphological word derivation        |         |      |         | 0.45    | 0.27                              | 0.17    | 0.40    | 0.27                              | 0.15    | 0.31    | 0.26                              | 0.11    |
| Arabic morphological choice                 |         |      |         | 0.03    | 0.10                              | 0.04    | 0.02    | 0.10                              | 0.02    | 0.01    | 0.09                              | 0.02    |
| Arabic reading comprehension                |         |      |         | 0.12    | 0.07                              | 0.19    | 0.07    | 0.07                              | 0.12    | 0.02    | 0.07                              | 0.03    |
| Hebrew morphological real word derivation   |         |      |         |         |                                   |         | 0.05    | 0.12                              | 0.04    | -0.01   | 0.11                              | -0.01   |
| Hebrew morphological pseudo word derivation |         |      |         |         |                                   |         | 0.12    | 0.07                              | 0.198   | 0.07    | 0.07                              | 0.12    |
| English morphological awareness             |         |      |         |         |                                   |         |         |                                   |         | 0.16    | 0.06                              | 0.33**  |
| English reading comprehension               |         |      |         |         |                                   |         |         |                                   |         | 0.03    | 0.06                              | 0.06    |
| $R^2$                                       |         | 0.22 |         |         | 0.30                              |         |         | 0.34                              |         |         | 0.42                              |         |
| $R^2$ change                                |         | 0.22 |         |         | 0.08                              |         |         | 0.04                              |         |         | 0.09                              |         |
| F for change in $R^2$                       |         |      |         |         | $F_{(3,76)} = 3.02$<br>$p = 0.04$ |         |         | $F_{(2,74)} = 2.03$<br>$p = 0.14$ |         |         | $F_{(2,72)} = 5.25$<br>$p = 0.01$ |         |

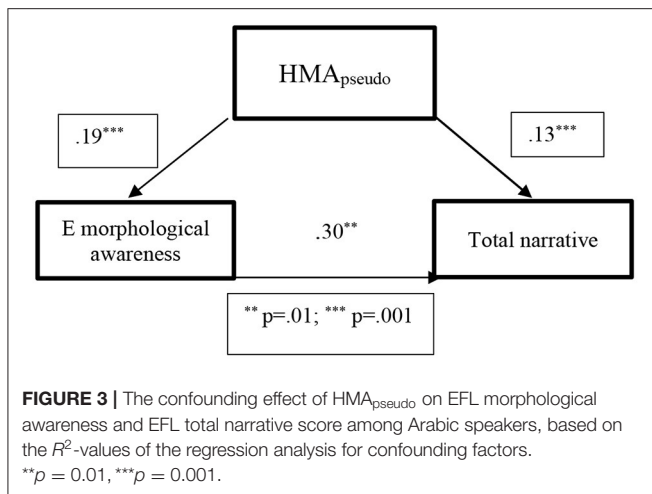
\*\* $p$  is significant at 0.01; \*\*\* $p$  is significant at 0.001.



play a role in that relationship. That skill (or variable) should have an influence on both, the predictor variable and outcome variable. This third variable should relate to both factors and possibly enhance the relationship between them. If we identify the predictor variable as X and the outcome variable as Y, variable Z (confounding variable) should have an effect on both, X and Y, therefore signifying cross linguistic transfer from L1(L2/L3)

to EFL. These relationships are represented in **Figure 1**. In what follows, we present the results for the regression analysis for confounding factors.

Among the Hebrew speakers, this regression analysis indicated that ERC was the strongest EFL predictor for total oral narrative scores [ $R^2 = 0.28$ ,  $F_{(1,84)} = 31.026$ ,  $p \leq 0.001$ ]. Contrary to the possible assumption that HRC may have some influence on the predictor and outcome variables based on the significant relationship between these variables found in the correlational analysis, this was not the case for Hebrew speakers. While there was strong correlation between ERC and HRC, no correlation was found between the Total Narrative score and HRC, therefore, violating the assumption that the confounding variable should have an influence on both the predictor and the outcome variable (MacKinnon et al., 2000). However, HMA<sub>pseudo</sub> was found to relate to both, ERC and Total Narrative score. The fact that HMA<sub>pseudo</sub> can be the influential variable in EFL Total Narrative score is supported by numerous studies, that suggest strong relationship between MA and RC and MA and oral language development (Tomasello, 2005; Verhoeven, 2017). The choice of HMA<sub>pseudo</sub> as possible variable that may enhance the strength of the relationship between EFL skills and EFL narratives was dictated by the strong metalinguistic component of this particular task: manipulation of pseudo words implies strong knowledge of the underlying mechanism in creating appropriate morphosyntactic elements (Shahar Yames et al., 2018). Therefore, based on the assumption of the confounding hypothesis regarding the effect of a third variable in enhancing the relationship between the predictor and outcome variable, we wanted to see if HMA<sub>pseudo</sub> would have an effect on both, the ERC and Total narrative score. Indeed, the effect of HMA<sub>pseudo</sub> on ERC was significant, suggesting that it would increase the ERC (predictor variable) score by 12% [ $R^2 = 0.122$ ,  $F_{(1,84)} = 11.48$ ,  $p < 0.001$ ]. The effect of the HMA<sub>pseudo</sub> on Total Narrative (outcome variable) score was also significant and suggested the increase on the Total Narrative score of 7% [ $R^2 = 0.073$ ,  $F_{(1,84)} = 6.601$ ,  $p < 0.01$ ]. Thus, we could say that HMA<sub>pseudo</sub>



had a significant effect on both, the predictor and the outcome variables, which was confirmed by bootstrapping procedures showing a confidence interval range from 0.01 to 0.13 and Sobel test ( $z = 2.08$ ,  $p = 0.04$ ) (See **Figure 2**).

Among the Arabic speakers, the regression analysis for confounding factors indicated that English MA was the strongest predictor for the Total oral narrative scores [ $R^2 = 0.302$ ,  $F_{(1,82)} = 35.445$ ,  $p \leq 0.001$ ], accounting for 30% of the variance. Again, our assumption that AMA may support EFL narratives did not yield the expected results. However, among Arabic speakers,  $AMA_{choice}$  and  $HMA_{pseudo}$  variables showed the strongest and most significant correlations. Moreover, significant correlations among this group were also found between  $HMA_{pseudo}$  and EMA, as well as Total Narrative score. As stated above, the relatively stronger performance of Arabic speakers in the task of  $HMA_{pseudo}$ , as opposed to  $HMA_{real}$  signified their developed metacognitive and metalinguistic skills, based on the exposure to multiple languages (Jessner, 2008). Therefore, we assumed that  $HMA_{pseudo}$  may have an effect on EFL Total Narrative score among this group as well. Indeed,  $HMA_{pseudo}$  had a significant effect on EMA among Arabic speakers, accounting for 19% of the variance [ $R^2 = 0.194$ ,  $F_{(1,82)} = 19.55$ ,  $p < 0.001$ ]. The effect of  $HMA_{pseudo}$  on the Total Narrative score in this group was even stronger than among Hebrew speakers, accounting for almost 13% of the variance [ $R^2 = 0.129$ ,  $F_{(1,82)} = 12.003$ ,  $p < 0.001$ ] and suggesting the cross language influence. Again, the bootstrapping procedures confirmed the significance of the effect, with a confidence interval range from 0.023 to 0.177, as did the Sobel test ( $z = 0.2.27$ ,  $p = 0.02$ ). Therefore, we assumed that  $HMA_{pseudo}$  was an influential variable that significantly impacted narrative performance in this group of participants (see **Figure 3**).

## DISCUSSION

The present study explored cross linguistic cognitive and linguistic Hebrew, MSA, and EFL influences on oral narrative production in EFL among two different Semitic L1 groups: Hebrew and Arabic speakers. We chose to focus our inquiry

on oral narrative skills for several reasons: (1) Oral narratives have been identified as a multidimensional measure of language proficiency in any language, as they encompass linguistic skills, i.e., lexical diversity and morphosyntactic knowledge, along with metacognitive abilities and global understanding of narrative structure; (2) Prior research has indicated strong relations between lexical and grammatical knowledge across different languages (Thordardottir et al., 2002), as well as across multiple language speakers, when producing narratives in non-native languages (Marchman et al., 2004; Gutiérrez-Clellen and Simon-Cerejido, 2009; Gagarina et al., 2015; Gagarina, 2016); and (3) Strong cross-linguistic influences have been seen in acquisition and use of morphological and morphosyntactic knowledge among speakers of different languages in their attempts to learn another language (Castilla et al., 2009; Zhang, 2015). Moreover, these influences show bidirectionality (Rezzonico et al., 2016).

It is important to note that while both of Hebrew and Arabic—the two Semitic languages under consideration in the current study—share many typological characteristics, there are also some features that are unique to Arabic resulting from its diglossic nature. Firstly, the distance between the spoken dialects and the MSA form amplifies the overall linguistic morphological complexity. Secondly, the formal acquisition of the MSA form, which coincides with the entry of an Arabic speaking child into school, has been found to hinder the acquisition of literacy and literacy related skills in both the early and later years (Saiegh-Haddad, 2003, 2007; Abdelhadi et al., 2011). Finally, in Israel, the MSA form, essentially an L2 for all Arabic-speaking children, is quickly followed by Hebrew, as an L3 and then EFL, as an L4. In essence, an Arabic speaking child is bilingual before he starts learning Hebrew and then EFL (Eviatar and Ibrahim, 2000), as opposed to a monolingual Hebrew speaking child who acquires EFL as his first additional language (L2). Within this context, it was of particular interest to investigate possible within group relational differences between L1 (Hebrew)/L2 (MSA), and EFL language skills among Arabic and Hebrew speakers for two reasons, (1) while Hebrew and Arabic are both Semitic languages, there are unique linguistic features in Arabic which could lead to different relationships between L2 (MSA) and EFL, as opposed to L1 Hebrew and EFL; (2) despite these differences in linguistic backgrounds, both language groups learn EFL according to the same curricular materials, and are subject to the same national exams.

Within this linguistic setting, our investigation was rooted in specific theoretical assumptions regarding cross-linguistic transfer, and the role played by language typology and L1 proficiency. In line with the *Linguistic Proximity Model* (Westergaard et al., 2017), specific patterns of influence will be determined by the areas of cross language overlap, or structural similarities among the languages of a multilingual learner. Therefore, we assumed that L1 (Hebrew) and L2 (MSA) skills would show shared underlying language learning mechanisms, based on typological proximity. However, we also acknowledged the potential role of the ambient language proficiency in the acquisition of additional languages, as postulated by the *Interdependence Hypothesis* (Cummins, 1979). In the present study, we were interested in exploring if these hypotheses are relevant when considering EFL oral language skills, as measured

by the production of oral narratives, in response to a static stimulus. Of specific interest were the issues of cross linguistic influences, as well as identifying modulating and mediating roles of specific EFL and native language skills in producing oral narratives in EFL. Addressing these issues would shed the light on the underlying mechanisms for CLIs between typologically distanced languages.

## The Influence of Cognitive and Native Language Skills Among Hebrew and Arabic Speakers on EFL Oral Narratives Production

Our first and second research questions addressed the influence of cognitive skills, Hebrew and MSA, as well as Hebrew as L3 for Arabic speakers, on all aspects of EFL narrative production. Our hypotheses stated that we should see cross linguistic influences of native language skills on EFL narratives based on levels of language proficiency, as postulated by the *Interdependence Hypothesis* (Cummins, 1979, 2014). Absence of the observed transfers could be attributed to the tenet of the *Linguistic Proximity Model* (Westergaard et al., 2017), suggesting that based on the typological distances between language non-facilitative influences may occur when the learner does not have a solid grasp of particular linguistic input in the target language. Based on these theoretical assumptions, we first looked at the overall psycholinguistic profiles of both linguistic groups in our study, beginning with the Hebrew speaking sample.

As seen in the descriptive statistics, Hebrew speakers showed high scores on measures of L1 proficiency, with higher scores for morphological awareness of real over pseudo words. In EFL, scores on morphological awareness (EMA) were at a moderate high level of proficiency. Scores for reading comprehension (ERC) were the lowest of the English linguistic scores, possibly indicating that despite strong knowledge of individual words, these children still lack the ability to recruit their vocabulary skills in order to comprehend language in context (Masrai, 2019). The scores for EWR as a measure of PM, were close to ceiling. Of interest here is the suggestion proposed by Kaushanskaya et al. (2011) that bilingual individuals may rely on PM for vocabulary retrieval in L2, therefore boosting reading comprehension as well as oral narratives. Furthermore, the fact that the PM measure (EWR) correlated with Hebrew morphology tasks supports the notion that PM is important not only for vocabulary building, but also for the acquisition of morphology independent of native language (Williams and Lovatt, 2003).

Our subsequent analyses, related to our first research question, aimed to identify the specific skills that show cross linguistic influences from L1 to EFL, and to see how these skills support oral language narrative production in EFL. Of interest was the finding that whereas Hebrew morphological awareness tasks (HMA) (real and pseudo words), as part of within language associations, did not show any correlations with Hebrew reading comprehension (HRC), both HMA tasks significantly correlated with EFL reading comprehension. It may be inferred, then, that strong correlations between HMA and ERC suggest that skills in this area of L1 linguistic knowledge are very important for

understanding the syntactic structure of L2 (EFL in this instance) in order to fully comprehend written text (Chen and Schwartz, 2018). This correlation exemplifies a cross-linguistic influence, and provides additional support to previous research, which indicated that morphological awareness could transfer from L1 to L2 even among typologically distant languages (Geva, 2006; Yip and Matthews, 2007; Pasquarella et al., 2011). Since HMA also showed strong association with EFL micro-, macro- and total score elements of EFL narratives, we can assume that L1 morphological awareness, as part of morphosyntactic knowledge (James et al., 2021) is a strong component of EFL oral production. Not surprisingly, the EWR as a measure of PM, showed strong associations with both HMA tasks, as well as all EFL narrative structures scores. This particular association provides strong support for the notion that PM is an important underlying cognitive/linguistic skill required for language acquisition. This relationship is seen not only across different L1 languages, but also in acquiring foreign languages as well (Service and Kononen, 1995; Masoura and Gathercole, 2005; Verhagen and Leseman, 2016). Moreover, the regression analysis highlighted the contribution of PM to the production of EFL narratives, solidifying our assumption that PM is important for acquisition of EFL oral language skills. Since Hebrew speakers showed high language proficiency in their ambient language, it supports the main tenet of *Interdependence Hypothesis* (Cummins, 1979, 2014). And yet, despite strong associations between the HMA tasks and all elements of EFL narratives, none of the Hebrew linguistic tasks added to the quality of EFL narratives beyond the contribution of EWR as a measure of PM.

Our second research question addressed the relationship between MSA (L2), L3 (Hebrew, for Arabic speakers), and EFL narratives among the Arabic speakers in the current study. It implied that there may be unique allocation of cognitive and linguistic resources among multilingual Arabic speakers. Our results indicated that our hypothesis was correct, as the results showed different within- and across-languages patterns of associations between tasks, as well as different cognitive and linguistic contributors to oral EFL narrative skills from those of the Hebrew speakers.

As a group, Arabic speakers showed average native language proficiency (according to existing norms, Asadi et al., 2015) and EFL proficiency within the average range, as measured by EMA and ERC scores. While the EFL tasks were not standardized as they were designed for the present study, achieving a result within the 33<sup>rd</sup>–50<sup>th</sup> percentile can be considered within average range<sup>3</sup>. This was not surprising given the diglossic nature of Arabic and the notion that the existence of two forms of Arabic (the spoken dialect and MSA form) may interfere in acquisitional

<sup>3</sup>On a teacher made or non-standardized test, as was the case with our experimental measures, raw scores represent the number of correct responses out of x number of questions. In our case percent correct was used to judge the child's ability on specific task. In line with what is generally accepted in psychological/psycholinguistic testing, any number above 50–60 percentile will be considered above average. Any number above 75 percentile will be high average. Subsequently, any number between 35 and 50 percentile will be within average range, below 35 percentile below average and below 16 percentile low score (Rumsey, 2011).

processes (Saiegh-Haddad, 2003, 2007). Moreover, the Arabic speakers in this study also had to contend with Hebrew as their third spoken and second written language. However, Arabic speakers showed strong results on EWR task as a measure of PM, as could be expected from individuals who are exposed to multiple phonological systems (Bialystok et al., 2003). This group also showed different pattern of results on the HMA tasks. In contrast to the Hebrew speakers, the Arabic speakers scored higher on the pseudo word derivation task than on the real word derivation task. This could be due to the fact that deriving a real word is more reliant on lexical knowledge, whereas deriving a pseudo word is more dependent on the underlying morpho-syntactic knowledge and internalization of rules and patterns (Williams and Lovatt, 2003). Thus, their scores would indicate that while their word level knowledge in Hebrew may be lower, their understanding of how words are formed may be heightened as a result of their experiences with multiple languages (Kuo and Anderson, 2006). Nevertheless, we also saw the same inter-language relationships among Arabic speakers in MSA measures, general cognitive skills and EWR, as a measure of PM, as were found among Hebrew speakers again indicating the important role of PM in language acquisition across languages.

Correlational analysis showed very strong associations between Raven (general cognition) and ARC, and every EFL measure, which was not the case for Hebrew speakers. Therefore, we can postulate that general cognitive abilities play an important role in language acquisition among this group of children. The reliance on general cognition for linguistic tasks has been postulated to be a driving force in language development (Clark, 2004; Tomasello, 2005). Tomasello (2005) specifically identified “pattern-finding...” as a “...cognitive skill involved in the abstraction process” (p. 193), which leads to integration of perceptual information into children’s linguistic repertoire (Clark, 2004). However, Raven also showed strong correlations with ARC and HMA for real words (HMA<sub>real</sub>). One possible explanation could suggest that Arabic speakers were relying on general cognitive resources to retrieve vocabulary items for reading comprehension in their native language, as well as recalling Hebrew vocabulary words to apply derivational process by association, rather than relying only on PM, although previous research found no connection between Raven and vocabulary (Ordóñez et al., 2002). On the other hand, we also saw strong associations between PM and AMA and ARC, as well as with all aspects of narrative structure. This finding is not surprising since PM is strongly related to vocabulary development (Gathercole et al., 1999), and has been found to strongly associate with both MA and RC, a skill which is comprised of decoding abilities, lexical knowledge, as well as knowledge and understanding of morphosyntactic structures (Hipfner-Boucher et al., 2015). As Arabic morphology is very rich, PM provides the basis in the development of all of the abovementioned skills, and in turn may support the acquisition of MA in additional languages (Verhagen and Leseman, 2016). Indeed, there were significant correlations between HMA and AMA, as an example of cross-linguistic transfer between two typologically close languages. Additionally, there was a cross-linguistic transfer in MA between Arabic and

EFL, a typologically distant language. These results could be interpreted in relation to the *Interdependence Hypothesis*. As PM was one of the strongest skills exhibited by Arabic speakers, we may postulate that the exposure to more than one phonological system has strengthened PM skills in this group. This conclusion gains support from numerous studies showing that bilingual children have stronger PM (Bialystok et al., 2003; Parra et al., 2011; Zaretsky, 2018).

## The Effect of EFL Language Skills in EFL Oral Narrative Production

Our assumptions from the onset of the study were that EFL skills will directly contribute to the EFL oral narrative production. The question was: will there be differences in these contributions among Hebrew and Arabic speakers. To answer this question, we conducted an additional regression analysis for both groups. Indeed, there were between group differences in how and which EFL skills supported oral language narratives. Among Hebrew speakers ERC made the largest contribution to the Total narrative score (combination of micro- and macro-structures). This contribution is not unexpected, as RC involves many elements that comprise narratives, as well as being a combination of lexical and morphosyntactic knowledge on its own. Moreover, Hebrew speakers showed above average scores on ERC (while there are no standard norms for this task since it was designed for this study, achieving a score between 60<sup>th</sup> and 75<sup>th</sup> percentile can be considered above average. See footnote 3 above), as well as high average score on HRC. This result is in line with the *Interdependence Hypothesis*, which postulates that L1 proficiency can support L2 skills. The fact that PM also supported EFL narratives in this group, was an additional conformation that this cognitive/linguistic skill is intimately involved in every aspect of language acquisition, from early acquisition of vocabulary (Gathercole et al., 1999) to supporting MA and metalinguistic skills (Kupersmitt et al., 2014; Boerma et al., 2016), as well as being an important skill in the acquisition of narrative abilities and promoting grammatical competence in later stages of L2 acquisition (O’Brien et al., 2006).

As we hypothesized, Arabic speakers relied on different EFL skills to support EFL oral narratives. The multilingual Arabic speakers were strong on the tasks that measured AM. Moreover, based on their HMA performance, specifically on the HMA pseudo word task, we can infer that they have better understanding of morphological procedures, above and beyond specific lexical representation at the word level. This finding is a direct conformation of the proposal presented by Shahar Yames et al. (2018), that suggested a separation of lexical and morphological knowledge among learners of additional language, who show much better performance on morphological tasks that are not dependent on extensive lexical knowledge, as measured in their case by pseudo word tasks. As further evidence of strong morphological awareness skills, the Arabic speakers performed within average range on EMA task (see footnote 3). Thus, it was no surprise to see that EMA was the largest contributor to the EFL Total narrative score. This particular finding was of importance as it provided additional support to previous

research findings that MA is essential for lexical development (Zhang, 2015), which in turn impacts oral narrative production. Moreover, the correlations between MA tasks in Arabic, Hebrew and EFL support the idea that cross-linguistic influences can be traced between languages that belong to the same typological group, such as Hebrew and Arabic, as well as languages that are typologically distant (Schwartz et al., 2016; Asli-Badarneh and Leikin, 2019), and further highlights the potential for cross linguistic influences across all the languages in the linguistic repertoire for multilinguals (Cenoz, 2013).

These results also highlighted skills that can be considered as contributors which serve to increase the level of oral language production among Hebrew and Arabic speaking children. It has been empirically shown that bilinguals and multilinguals are influenced by one of their languages in activating the processing of another (Dijkstra, 2005), and this activation is seen for different types of information, including syntactic structures (Macizo et al., 2010). So, it is possible to postulate that ERC is the important influencer for EFL oral narratives for Hebrew speakers, i.e., the stronger the ERC scores, the better the EFL narratives, especially since RC and narratives rely on similar component skills. Further, ERC scores for this group were very strong. However, there was a different influence of EFL narrative scores for Arabic speakers, namely EMA. Influence of MA on narratives skills are well-documented, therefore there is no surprise that there should be an influence of EMA on EFL narratives, particularly in this group. This relationship in this case may be explained by the fact that the Arabic speakers may have heightened morphological awareness skills as part of their cross linguistic metacognitive repertoire (Bialystok et al., 2003; Hirosh and Degani, 2018), resulting from their prior experiences with two Arabic morphological systems (spoken dialect and MSA form), as well as Hebrew MA. This could explain how increases in EMA may reflect an increase in the EFL Total narrative score among this population.

### **L1(Hebrew)/L2 (Modern Standard Arabic)/L3 (Hebrew for Arabic Speakers): The Skills That Play an Influential Role in the Relationship Between EFL Skills and EFL Oral Narratives**

Our last research question explored the possibility that there may be a specific L1/L2/L3 skill, that could play an influencing role in enhancing the relationship between the predictor and outcome variables. We chose the confounding hypothesis as an explanatory framework in order to explore this possibility. The confounding hypothesis proposes a relationship whereby an additional variable should be related to the factors of interest (predictor and outcome). Namely, it should be correlated with predictor and related to the outcome. Moreover, this particular hypothesis would strongly support the tenets of *Interdependence Hypothesis*.

As seen in the strength of the EFL skills as predictors for EFL narratives, ERC was the predictor variable for Hebrew speakers. The correlational analysis suggested that HMA<sub>pseudo</sub> (as the third variable) strongly correlated with ERC, our independent

variable (IV, or predictor), therefore fulfilling the first assumption of the confounding hypothesis, namely that the confounding variable and IV should correlate. At the same time, HMA<sub>pseudo</sub> was related to the EFL Total Narrative score, i.e., it increased the quality of the narrative, fulfilling the second assumption of the confounding hypothesis. Moreover, this explanation aligns with the tenet of the *Interdependence Hypothesis*, in suggesting that strong L1 skills will support L2 skills, indicating a CLI.

Among Arabic speakers, EMA was the predictor variable for the Total narrative score, as evidence of stronger metalinguistic knowledge in this multilingual group as opposed to weaker specific linguistic knowledge. However, AMA was not found to explain the relationship between the predictor (EMA) and the outcome (EFL Total narrative score), e.g., it was not a factor in enhancing the outcome. On the other hand, the strong performance of Arabic speakers on HMA<sub>pseudo</sub> task also suggested the possibility that HMA may be the 3<sup>rd</sup> variable that would explain the relationship between IV (EMA) and DV (EFL Total narrative). Our analysis indicated that this was the case for Arabic speakers as well: HMA<sub>pseudo</sub> was a confounding variable that explained the increase in EFL oral language performance among Arabic speakers. This finding is important because it expands the significance of previous data regarding the role of morphological awareness as L1/L2 linguistic skills (Zhang, 2015), to include additional languages. Moreover, this particular finding provided strong support for previous research findings that MA is essential for vocabulary development (Zhang, 2015), and suggested that it can play a strong role in acquisitional processes.

The findings for both language groups are also in line with Hirosh and Degani (2018) proposition that multilinguals may activate both direct routes and indirect routes to additional language learning, where direct routes include transfer of linguistic skills and knowledge, while indirect routes represent cognitive factors such as metalinguistic awareness and working memory, which was the case among our participants. The present findings gain support from recent research which indicates the importance of crosslinguistic transfer of skills, particularly MA, from L1 to L2 (*forward transfer*) (Jarvis and Pavlenko, 2008; Kim and Piper, 2019). Moreover, we found that a specific MA skill in one language may enhance narrative production among two groups of speakers of different Semitic languages. This skill was HMA<sub>pseudo</sub>, which clearly influenced the EFL Total narrative scores for both linguistic groups, despite typological distance of the languages. It is possible then to postulate that the assumptions of *Linguistic Proximity Model* (Westergaard et al., 2017) that crosslinguistic transfer between typologically distant languages will be determined by the areas of cross language overlap may not necessarily account for the application of all cognitive and linguistic resources among multilingual speakers, since our findings indicated transfer among typologically different languages. Although we did not observe bidirectional transfer of specific skills, the finding that Hebrew MA support EFL narratives not only for native Hebrew speakers but also for Arabic speakers, increases our understanding of the role of CLIs in oral language production.

## CONCLUSIONS

The present study provided important insights regarding the cognitive and linguistic skills contributing to oral narrative production in EFL among speakers of Semitic languages: Arabic and Hebrew. In line with the theoretical framework suggested by the *Interdependence Hypothesis* (Cummins, 1979, 1991), we saw cross-linguistic transfer of both cognitive and linguistic skills between Arabic and Hebrew, typologically close languages, as well as cross-linguistic transfer, particularly in morphological awareness, in typologically distant languages (Arabic, Hebrew, and English), although the specific psycholinguistic cross linguistic profiles of each language group were unique. Nonetheless, Hebrew morphological derivation of pseudo words was found to be a confounding variable for total narrative skills in both language groups, thereby adding support to previous findings that MA is an essential skill required for native language acquisition, as well as for acquisition of an additional language (Kuo and Anderson, 2006).

The findings of this study could also have pedagogical repercussions especially in light of the fact that in today's world many pupils study in languages that they do not speak at home (Nag et al., 2019) and specifically in Israel, where pupils from multiple language backgrounds all study English according to the same curriculum. In line with the *Linguistic Proximity Model* (Westergaard et al., 2017), our analyses suggested that even typologically related linguistic groups may exhibit different allocations of cognitive and linguistic resources to achieve L1 and L2 and EFL proficiency. Thus, if pupils from different L1 backgrounds are expected to study according to the same curricular materials and be tested according to the same standards, it is possible that certain implementational modifications in the study programs should be made based on the areas of linguistic overlap and the breadth of the linguistic repertoires of the languages of the learners. This could be as simple as acknowledging fine-grained similarities and differences between languages in a direct manner during the teaching process. Moreover, it is possible that for multilingual learners who have not reached a sufficient level of proficiency to support

cross linguistic transfer across languages, it may be prudent to allocate additional hours for extended practice, specifically in situations where English is the third or fourth language of the learners. These implications are particularly important since previous research has highlighted the need for specific and targeted intervention in order to aid EFL learners in acquiring necessary proficiency in English across oral and written modalities (Kahn-Horwitz, 2020).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Chief Scientist of the State of Israel. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

The data collection for this study was supported by the Israel Science Foundation, grant number 592/14. Partial funds to cover open access publication fees are being provided by Beit Berl College Israel.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.664152/full#supplementary-material>

## REFERENCES

- Abdelhadi, S., Ibrahim, R., and Eviatar, Z. (2011). Perceptual load in the reading of Arabic: effects of orthographic visual complexity on detection. *Writ. Syst. Res.* 3, 117–127. doi: 10.1093/wsr/wsr014
- Al Hosni, S. (2014). Speaking difficulties encountered by young EFL learners. *Int. J. Stud. Engl. Lang. Liter.* 2, 22–30.
- Amara, M. H. (2018). Palestinian schoolscapes in Israel. *Asian Pac. J. Second Foreign Lang. Educ.* 3:7. doi: 10.1186/s40862-018-0047-1
- Asadi, I. M., Shany, A., and Ben-Simon, and Ibrahim, R. (2015). *Individual Diagnostic Tests in the Assessment of Learning Disabilities in Arabic: Tests and Manual*. Haifa: Haifa University Publication (in Arabic).
- Asli-Badarnah, A., and Leikin, M. (2019). Morphological ability among monolingual and bilingual speakers in early childhood: the case of two Semitic languages. *Int. J. Bilingual.* 23, 1087–1105. doi: 10.1177/1367006918781079
- Berman, R. A. (1995). Narrative competence and storytelling performance: how children tell stories in different contexts. *J. Narrat. Life Hist.* 5, 285–313. doi: 10.1075/jnlh.5.4.01nar
- Berman, R. A., and Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bialystok, E., Majumde, S., and Martin, M. (2003). Developing phonological awareness: is there a bilingual advantage? *Appl. Psycholinguist.* 24, 27–44. doi: 10.1017/S014271640300002X
- Boerma, T., Leseman, P., Timmermeister, M., Wijnen, F., and Blom, E. (2016). Narrative abilities of monolingual and bilingual children with and without language impairment: implications for clinical practice. *Int. J. Lang. Commun. Disord.* 51, 626–638. doi: 10.1111/1460-6984.12234
- Castilla, A. P., Restrepo, M. A., and Perez-Leroux, A. T. (2009). Individual differences and language interdependence: a study of sequential bilingual development in Spanish–English preschool children. *Biling. Biling. Educ.* 12, 565–580. doi: 10.1080/13670050802357795
- Cenoz, J. (2013). The influence of bilingualism on third language acquisition: focus on multilingualism. *Lang. Teach.* 46, 71–86. doi: 10.1017/S0261444811000218
- Chen, X., and Schwartz, M. (2018). Morphological awareness and literacy in second language learners: a cross-language

- perspective. *Read. Writ.* 31, 1685–1694. doi: 10.1007/s11145-018-9866-1
- Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends Cogn. Sci.* 8, 472–478. doi: 10.1016/j.tics.2004.08.012
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104
- Cornaglia, A., Brandel, J., and Hoffman, L. (2017). “Narrative quality and elicitation procedures: is there a relationship?,” *Poster Presentation at AHSA Convention* (Washington, DC).
- Cortazzi, M., and Jin, L. (2007). Narrative learning, EAL and metacognitive development. *Early Child Dev. Care* 177, 645–660. doi: 10.1080/03004430701379074
- Cromley, J. G., and Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *J. Educ. Psychol.* 99:311. doi: 10.1037/0022-0663.99.2.311
- Crystal, D. (2012). *English as a Global Language, 2nd Edn. Canto Classics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139196970
- Cummings, L. (2019). Describing the Cookie Theft picture: sources of breakdown in Alzheimer's dementia. *Prag. Soc.* 10, 153–176. doi: 10.1075/ps.17011.cum
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Rev. Educ. Res.* 49, 222–251. doi: 10.3102/00346543049002222
- Cummins, J. (1991). “Interdependence of first- and second-language proficiency in bilingual children,” in *Language Processing in Bilingual Children*, ed E. Bialystok (Cambridge: Cambridge University Press), p. 70–89. doi: 10.1017/CBO9780511620652.006
- Cummins, J. (2014). Beyond language: academic communication and student success. *Linguist. Educ.* 26, 145–154. doi: 10.1016/j.linged.2014.01.006
- DeKeyser, R., Alfi-Shabtay, I., and Ravid, D. (2010). Crosslinguistic evidence for the nature of age effects in second language acquisition. *Appl. Psycholinguist.* 31, 413–438. doi: 10.1017/S0142716410000056
- Dickinson, D. K., Hofer, K. G., and Rivera, B. L. (2019). “The developing language foundation for reading comprehension,” in *Narrative, Literacy and Other Skills: Studies in Intervention*, Vol. 25, eds E. Veneziano and A. Nicolopoulou (Amsterdam: John Benjamins Publishing Company), 41–61. doi: 10.1075/sin.25.02dic
- Dijkstra, T. (2005). “Bilingual word recognition and lexical access,” in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. De Groot (New York, NY: Oxford University Press), 179–201.
- El Akiki, C., and Content, A. (2020). Early sensitivity to morphology in beginning readers of Arabic. *Front. Psychol.* 11:552315. doi: 10.3389/fpsyg.2020.552315
- Eviatar, Z., and Ibrahim, R. (2000). Bilingual is as bilingual does: metalinguistic abilities of Arabic-speaking children. *Appl. Psycholinguist.* 21, 451–471. doi: 10.1017/S0142716400004021
- Friend, M., and Bates, R. P. (2014). The union of narrative and executive function: different but complimentary. *Front. Psychol.* 5:469. doi: 10.3389/fpsyg.2014.00469
- Gagarina, N. (2016). Narratives of Russian-German preschool and primary school bilinguals: Rasskaz and Erzählung. *Appl. Psycholinguist.* 37, 91–122. doi: 10.1017/S0142716415000430
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiuniene, I., et al. (2015). “Assessment of narrative abilities in bilingual children,” in *Methods for Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*, eds S. Armon-Lotem, J. de Jong and N. Meir (Bristol, UK: Multilingual Matters), 243–276. doi: 10.21832/9781783093137-011
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., and Martin, A. J. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Appl. Cogn. Psychol.* 13, 65–77. doi: 10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O
- Geva, E. (2006). “Second-language oral proficiency and second-language literacy,” in *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*, eds D. August and T. Shanahan (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 123–139.
- Gillis, S., and Ravid, D. (2006). Typological effects on spelling development: a crosslinguistic study of Hebrew and Dutch. *J. Child Lang.* 33, 621–659. doi: 10.1017/S0305000906007434
- Goodglass, H., and Kaplan, E. (1983). *The Assessment of Aphasia and Related Disorders*, 2nd Edn. Philadelphia, PA: Lea and Febiger.
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459
- Gutiérrez-Clellen, V. F., and Simon-Cerejido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Semin. Speech Lang.* 30, 234–245. doi: 10.1055/s-0029-1241722
- Hayes, A. F. (2019). Introduction to Mediation, Moderation, and Conditional Process Analysis, 2nd Edn., A Regression-Based Approach. New York, NY: Guilford Press.
- Heilman, J., Miller, J., Nockerts, A., and Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *Am. J. Speech Lang. Pathol.* 19, 154–166. doi: 10.1044/1058-0360(2009/08-0024)
- Hipfner-Boucher, K., Lam, K., and Chen, X. (2015). The contribution of narrative morphosyntactic quality to reading comprehension in French immersion students. *Appl. Psycholinguist.* 36, 1375–1391. doi: 10.1017/S0142716414000319
- Hirosh, Z., and Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: an integrative review. *Psychon. Bull. Rev.* 25, 892–916. doi: 10.3758/s13423-017-1315-7
- Huang, K. J. (2018). On bilinguals' development of metalinguistic awareness and its transfer to L3 learning: the role of language characteristics. *Int. J. Bilingual.* 22, 330–349. doi: 10.1177/1367006916681081
- Iluz-Cohen, P., and Walters, J. (2012). Telling stories in two languages: narratives of bilingual preschool children with typical and impaired language. *Bilingualism* 15, 58–74. doi: 10.1017/S1366728911000538
- James, E., Currie, N. K., Tong, S. X., and Cain, K. (2021). The relations between morphological awareness and reading comprehension in beginner readers to young adolescents. *J. Res. Read.* 110–130. doi: 10.1111/1467-9817.12316
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York, NY; London: Routledge. doi: 10.4324/9780203935927
- Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *Mod. Lang. J.* 92, 270–283. doi: 10.1111/j.1540-4781.2008.00718.x
- Justice, L. M., Bowles, R., Pence, K., and Gosse, C. (2010). A scalable tool for assessment children's language abilities within a narrative context: the NAP (Narrative Assessment Protocol). *Early Child. Res. Q.* 25, 218–234. doi: 10.1016/j.ecresq.2009.11.002
- Kahn-Horwitz, J. (2020). “I didn't even know one of the conventions before”: explicit EFL spelling instruction and individual differences. *Cogn. Dev.* 55:100880. doi: 10.1016/j.cogdev.2020.100880
- Kaushanskaya, M., Blumenfeld, H. K., and Marian, V. (2011). The relationship between short-term memory and vocabulary measures in monolingual and bilingual speakers. *Int. J. Bilingual.* 15, 408–425. doi: 10.1177/1367006911403201
- Kim, Y. S. G., and Piper, B. (2019). Cross-language transfer of reading skills: an empirical investigation of bidirectionality and the influence of instructional environments. *Read. Writ.* 32, 839–871. doi: 10.1007/s11145-018-9889-7
- Kormos, J., and Trebits, A. (2011). “Working memory capacity and narrative task performance,” in *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, ed P. Robinson (John Benjamins), 267–289. doi: 10.1075/tblt.2.17ch10
- Kuo, L. J., and Anderson, R. C. (2006). Morphological awareness and learning to read: a cross-language perspective. *Educ. Psychol.* 41, 161–180. doi: 10.1207/s15326985ep4103\_3
- Kupersmitt, J., Yifat, R., and Blum-Kulka, S. (2014). The development of cohesion and coherence in monolingual and sequential bilingual children's narratives: same or different? *Narr. Inq.* 24, 40–76. doi: 10.1075/ni.24.1.03kup
- Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Oxford: Blackwell.
- Leikin, M., Ibrahim, R., and Eghbaria, H. (2014). The influence of diglossia in Arabic narrative ability: evidence from analysis of the linguistic and narrative structure of discourse among pre-school children. *Read. Writ.* 27, 733–747. doi: 10.1007/s11145-013-9462-3
- Lucero, A. (2018). The development of bilingual narrative retelling among spanish-English dual language learners over two years. *Lang. Speech Hear. Serv. Sch.* 49, 607–621. doi: 10.1044/2018\_LSHSS-17-0152
- Luo, Y. C., Chen, X., and Geva, E. (2014). Concurrent and longitudinal cross-linguistic transfer of phonological awareness and morphological

- awareness in Chinese-English bilingual children. *Writ. Lang. Lit.* 17, 89–115. doi: 10.1075/wll.17.1.05luo
- Macizo, P., Herrera, A., Paolieri, D., and Román, P. (2010). Is there cross-language modulation when bilinguals process number words? *Appl. Psycholinguist.* 31, 651–669. doi: 10.1017/S0142716410000184
- MacKinnon, D. P., Krull, J. L., and Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prev. Sci.* 1, 173–181.
- Marchman, V. A., Martinez-Sussmann, C., and Dale, P. S. (2004). The language-specific nature of grammatical development: evidence from bilingual language learners. *Dev. Sci.* 7, 212–224. doi: 10.1111/j.1467-7687.2004.00340.x
- Masoura, E. V., and Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in foreign language. *Memory* 13, 422–429. doi: 10.1080/09658210344000323
- Masrai, A. (2019). Vocabulary and reading comprehension revisited: evidence for high-, mid-, and low-frequency vocabulary knowledge. *Sage Open* 9, 1–13. doi: 10.1177/2158244019845182
- Moattarian, A. (2013). Bidirectional crosslinguistic influence in language learning: linguistic aspects and beyond. *Int. J. Linguist.* 5, 38–49. doi: 10.5296/ijlv.v5i4.3746
- Moonsamy, S., Jordaan, H., and Greenop, K. (2009). Cognitive processing and narrative discourse production in children with ADHD. *South Afr. J. Psychol.* 39, 326–335. doi: 10.1177/008124630903900307
- Nag, S., Vagh, S. B., Dulay, K. M., and Snowling, M. J. (2019). Home language, school language and children's literacy attainments: a systematic review of evidence from low- and middle-income countries. *Rev. Educ.* 7, 91–150. doi: 10.1002/rev3.3132
- O'Brien, I., Segalowitz, N., Collentine, J., and Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Appl. Psycholinguist.* 27, 377–402. doi: 10.1017/S0142716406060322
- Ordóñez, C. L., Carlo, M. S., Snow, C. E., and McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: which vocabulary skills transfer? *J. Educ. Psychol.* 94, 719–728. doi: 10.1037/0022-0663.94.4.719
- Orland-Barak, L., and Yinon, H. (2005). Different but similar: student teachers' perspectives on the use of L1 in Arab and Jewish EFL classroom settings. *Lang. Cult. Curric.* 18, 91–113. doi: 10.1080/07908310508668735
- Parra, M., Hoff, E., and Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *J. Exp. Psychol.* 108, 113–125. doi: 10.1016/j.jecp.2010.07.011
- Pasquarella, A., Chen, X., Lam, K., Luo, Y. C., and Ramirez, G. (2011). Cross-language transfer of morphological awareness in Chinese-English bilinguals. *J. Res. Read.* 34, 23–42. doi: 10.1111/j.1467-9817.2010.01484.x
- Pavlenko, A. (2002). Bilingualism and emotions. *Multilingua* 21, 45–78. doi: 10.1515/mult.2002.004
- Pavlenko, A., and Jarvis, S. (2002). Bidirectional transfer. *Appl. Linguist.* 23, 190–214. doi: 10.1093/applin/23.2.190
- Pearson, B. Z. (2002). "Narrative competence among monolingual and bilingual school children in Miami," in *Language and Literacy in Bilingual Children*, eds D. K. Oller and R. E. Eilers (Clevedon, UK: Multilingual Matters), 135–174. doi: 10.21832/9781853595721-008
- Raven, J., Raven, J. C., and Court, J. H. (1976). *Raven's Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Ravid, D. (2001). Learning to spell in Hebrew: phonological and morphological factors. *Read. Writ.* 14, 459–485. doi: 10.1023/A:1011192806656
- Rezzonico, S., Goldberg, A., Mak, K. K. Y., Yap, S., Milburn, T., Belletti, A., et al. (2016). Narratives in two languages: storytelling of bilingual Cantonese-English preschoolers. *J. Speech Lang. Hear. Res.* 59, 521–532. doi: 10.1044/2015\_JSLHR-L-15-0052
- Rumsey, D. J. (2011). *Statistics for Dummies, 2nd Edn*. Hoboken, NJ: Wiley Publishing, Inc.
- Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: the case of Arabic diglossia. *Appl. Psycholinguist.* 24, 431–451. doi: 10.1017/S0142716403000225
- Saiegh-Haddad, E. (2007). Linguistic constraints of children's ability to isolate phonemes in Arabic. *Appl. Psycholinguist.* 28, 607–625. doi: 10.1017/S0142716407070336
- Saiegh-Haddad, E., and Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English-Arabic bilingual children. *Read. Writ.* 21, 481–504. doi: 10.1007/s11145-007-9074-x
- Saiegh-Haddad, E., and Henkin-Roitfarb, R. (2014). "The structure of Arabic language and orthography," in *Handbook of Arabic Literacy: Insights and Perspectives*, eds E. Saiegh-Haddad and M. Joshi (Dordrecht: Springer), 3–28. doi: 10.1007/978-94-017-8545-7\_1
- Saiegh-Haddad, E., and Taha, H. (2017). The role of morphological and phonological awareness in early development of word spelling and reading in typically developing and disabled Arabic readers. *Dyslexia* 23, 345–371. doi: 10.1002/dys.1572
- Schwartz, M., Geva, E., Share, D., and Leikin, M. (2007). Learning to read in English as a third language: the cross-linguistic transfer of phonological processing skills. *Writ. Lang. Lit.* 10, 25–52. doi: 10.1075/wll.10.1.03sch
- Schwartz, M., Taha, H., Assad, H., Khamaisi, F., and Eviatar, Z. (2016). The role of emergent bilingualism in the development of morphological awareness in Arabic and Hebrew. *J. Speech Lang. Hear. Res.* 59, 797–809. doi: 10.1044/2016\_JSLHR-L-14-0363
- Schwartzwald, O. (2011). "Modern Hebrew," in *The Semitic Languages: An International Handbook*, ed S. Weninger (Berlin: Mouton), 523–536.
- Service, E., and Kononen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Appl. Psycholinguist.* 16, 155–172. doi: 10.1017/S0142716400007062
- Shahar Yames, D., Eviatar, Z., and Prior, A. (2018). Separability of lexical and morphological knowledge: evidence from language minority children. *Front. Psychol.* 9:163. doi: 10.3389/fpsyg.2018.00163
- Shany, M., Lachman, D., Shalem, Z., Bahat, A., and Zeiger, T. (2006). "Aleph-Taph" A Test for the Diagnosis of Reading and Writing Disabilities, Based on National Israeli Norms. Tel Aviv: Yesod Publishing (in Hebrew).
- Simon-Cerejido, G., and Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Appl. Psycholinguist.* 30, 315–337. doi: 10.1017/S0142716409090134
- Soodla, P., and Kikas, E. (2010). Macrostructure in the narratives of Estonian children with typical development and language impairment. *J. Speech Lang. Hear. Res.* 53, 1321–1333. doi: 10.1044/1092-4388(2010/08-0113)
- State of Israel, Ministry of Education Pedagogical Secretariat, Language Department, Inspectorate for English Language Education. (2020). *English Curriculum for Elementary School*. Available online at: [https://meyda.education.gov.il/files/Mazkirut\\_Pedagogit/English/curriculum2020Elementary.pdf](https://meyda.education.gov.il/files/Mazkirut_Pedagogit/English/curriculum2020Elementary.pdf) (accessed July 2, 2021).
- Terry, N. P., Mills, T., Bingham, G. E., Mansour, S., and Marencin, N. (2013). Oral narrative performance of African American prekindergartners who speak nonmainstream American English. *Lang. Speech Hear. Serv. Sch.* 44, 291–305. doi: 10.1044/0161-1461(2013)12-0037
- Thordardottir, E. T., Weismer, S. E., and Evans, J. L. (2002). Continuity of lexical and morphological development in Icelandic and English-speaking 2-year-olds. *First Lang.* 22, 3–28. doi: 10.1177/014272370202206401
- Timor, T. (2012). Use of the mother tongue in teaching a foreign language education in Asia. *Lang. Educ. Asia* 3, 7–17. doi: 10.5746/LEiA/12/V3/I1/A02/Timor
- Tomasello, M. (2005). Beyond formalities: the case of language acquisition. *Linguist. Rev.* 22, 183–197. doi: 10.1515/tlr.2005.22.2-4.183
- Tsarfaty, R., and Sima'an, K. (2010). "Modeling morphosyntactic agreement in constituency-based parsing of Modern Hebrew," in *Proceeding of the NAACL HLT 2010 First Workshop on Statistically Parsing of Morphologically-Rich Languages* (Los Angeles, CA), 40–48.
- Uccelli, P., and Páez, M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Lang. Speech Hear. Serv. Sch.* 38, 225–236. doi: 10.1044/0161-1461(2007/024)
- Verhagen, J., and Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *J. Exp. Child Psychol.* 141, 65–82. doi: 10.1016/j.jecp.2015.06.015
- Verhoeven, L. (1994). Transfer in bilingual development: the linguistic interdependence hypothesis revisited. *Lang. Learn.* 44, 381–415. doi: 10.1111/j.1467-1770.1994.tb01112.x

- Verhoeven, L. (2017). "Ch. 13: Learning to read Dutch," in *Learning to Read Across Languages and Writing Systems*, eds L. Verhoeven and C. Perfetti (Cambridge University Press), 323–346.
- Westergaard, M., Mitrofanova, N., Mykhaylyk, R., and Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: the linguistic proximity model. *Int. J. Bilingual.* 21, 666–682. doi: 10.1177/1367006916648859
- Williams, J. N., and Lovatt, P. (2003). Phonological memory and rule learning. *Lang. Learn.* 53, 67–121. doi: 10.1111/1467-9922.00211
- Yip, V., and Matthews, S. (2007). Relative clauses in Cantonese-English bilingual children: typological challenges and processing motivations. *Stud. Second Lang. Acquisit.* 29, 277–300. doi: 10.1017/S0272263107070143
- Zaretsky, E. (2018). Verbal working memory and early literacy acquisition: do ELLs allocate resources similar to their monolingual peers or children with SLI? *Int. J. Biling. Educ. Bilingual.* 23, 1051–1070. doi: 10.1080/13670050.2018.1423672
- Zhang, H. (2015). Morphological awareness in vocabulary acquisition among Chinese-speaking children: testing partial mediation via lexical interference ability. *Read. Res. Quar.* 50, 129–142. doi: 10.1002/rrq.89

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Russak and Zaretsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Influence of Cross-Linguistic Similarity and Language Background on Writing to Dictation

Antonio Iniesta<sup>1\*</sup>, Eleonora Rossi<sup>2</sup>, M. Teresa Bajo<sup>1</sup> and Daniela Paolieri<sup>1</sup>

<sup>1</sup>Department of Experimental Psychology, Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain, <sup>2</sup>Department of Linguistics, University of Florida, Gainesville, FL, United States

## OPEN ACCESS

### Edited by:

Tamar Degani,  
University of Haifa, Israel

### Reviewed by:

Katherine I. Martin,  
Southern Illinois University  
Carbondale, United States  
Laura Muscali,  
Ithaca College, United States

### \*Correspondence:

Antonio Iniesta  
iniesta@ugr.es

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 12 March 2021

**Accepted:** 01 September 2021

**Published:** 28 September 2021

### Citation:

Iniesta A, Rossi E, Bajo MT and  
Paolieri D (2021) The Influence of  
Cross-Linguistic Similarity and  
Language Background on Writing to  
Dictation.  
Front. Psychol. 12:679956.  
doi: 10.3389/fpsyg.2021.679956

This study used a word dictation task to examine the influence of a variety of factors on word writing production: cognate status (cognate vs. non-cognate words), orthographic (OS) and phonological similarity (PS) within the set of cognate words, and language learning background [late bilinguals (LBs) with academic literacy and formal instruction in English and Spanish, and heritage speakers (HSs) with academic literacy and formal instruction only in English]. Both accuracy and reaction times for the first key pressed by participants (indicating lexical access), and the time required to type the rest of the word after the first keypress (indicating sublexical processing) was assessed. The results revealed an effect of PS on the dictation task particularly for the first keypress. That is, cognates with high PS were processed faster than cognates with low PS. In contrast to reading studies in which PS only revealed a significant effect when the OS between languages was high (O+P+ vs. O+P-), in the dictation to writing task, the phonology had a more general effect across all conditions, regardless of the level of OS. On the other hand, OS tended to be more influential for typing the rest of the word. This pattern is interpreted as indicating the importance of phonology (and PS in cognates) for initial lexical retrieval when the input is aural. In addition, the role of OS and PS during co-activation was different between groups probably due to the participants' linguistic learning environment. Concretely, HSs were found to show relatively lower OS effects, which is attributed to the greater emphasis on spoken language in their Spanish language learning experiences, compared to the formal education received by the LBs. Thus, the study demonstrates that PS can influence lexical processing of cognates, as long as the task demands specifically require phonological processing, and that variations in language learning experiences also modulate lexical processing in bilinguals.

**Keywords:** bilingual writing, language co-activation, orthographic/phonological similarity, heritage speakers, writing to dictation

## INTRODUCTION

A central question in bilingual research has been to determine how bilinguals manage the use of words from different languages (Kroll et al., 2013; Costa and Sebastián-Gallés, 2014). There is evidence that bilinguals co-activate their two languages, even in single language contexts (e.g., Van Heuven et al., 1998; Van Hell and Dijkstra, 2002; Von Studnitz and Green, 2002;

Marian and Spivey, 2003; Hoshino and Kroll, 2008; Macizo, 2016) and that this parallel co-activation may facilitate (Costa et al., 2000; Christoffels et al., 2007; Voga and Grainger, 2007; Lemhöfer et al., 2008) or hinder access to intended words (Gollan et al., 2005; Ivanova and Costa, 2008). Under the assumption that the two languages are co-activated (“non-selective” activation of the two languages; Dijkstra and Van Heuven, 2002), a key question is whether bilingual language co-activation is modulated at different linguistic levels (e.g., lexical, orthographic, and phonological) depending on the linguistic tasks (i.e., reading, speaking, and writing). Critically, one question that is untapped in the literature is how these various levels of co-activation and control thereof vary for different bilingual populations with diverse language experiences.

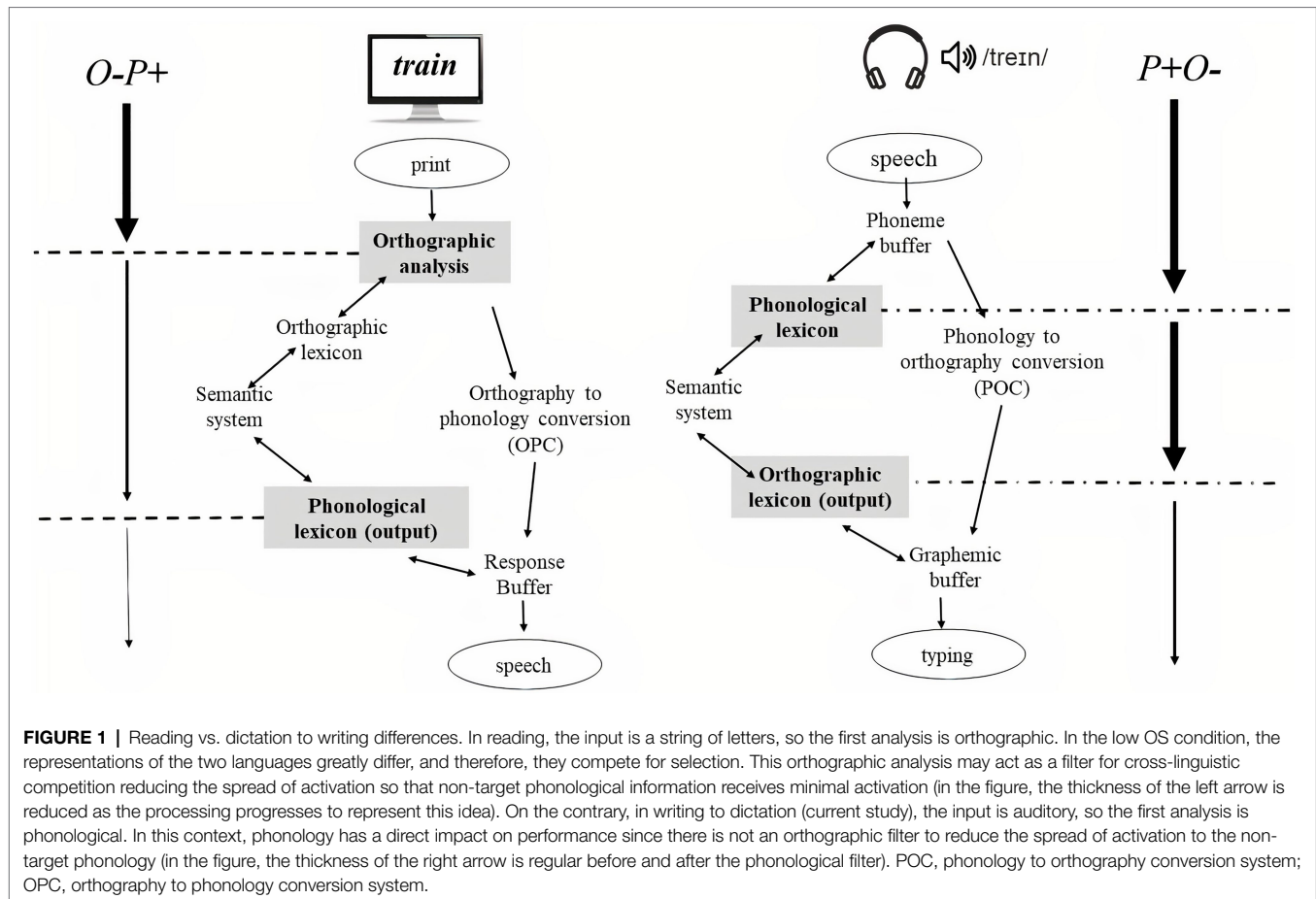
Orthographic processing has been the focus of most bilingual word recognition studies (e.g., van Heuven et al., 1998; Van Kesteren et al., 2012; Casaponsa et al., 2014; Hoversten et al., 2017). The cross-linguistic influence of the two bilingual orthographic codes has been strongly supported by experimental evidence using cognate words. Cognate words are words that have the same meaning and form representation in two or more languages (e.g., “chocolate” in English is translated as “chocolate” in Spanish). Behavioral studies using different experimental tasks (lexical decision, word recognition, naming, and translation) have demonstrated that cognate words are processed faster than non-cognates (words with different lexical representations between languages, i.e., “bed” in English and “cama” in Spanish). This evidence comes from studies in which the words were presented in the visual (e.g., Dijkstra et al., 1999; Costa et al., 2000; Hoshino and Kroll, 2008; Peeters et al., 2013) and the auditory modalities (Andras et al., under review; Woutersen et al., 1995; Bowers et al., 2000). Cognate facilitation has also been reported in spoken word production studies (Costa et al., 2005; see also Muscalu and Smiley, 2018 for typing). Thus, most models of bilingual language processing assume that both languages are co-activated and include predictions for the role of cognate words during word recognition (e.g., bilingual interactive activation BIA+ model; Dijkstra and Van Heuven, 2002) and word production (e.g., The revised hierarchical model – RHM; Kroll et al., 2010).

However, hypotheses regarding the processing of non-identical but similar cognates are not completely clear (Dijkstra et al., 2010). Cognate facilitation seems to be greater for identical cognates than non-identical cognates (Comesaña et al., 2015; Guasch et al., 2017) with larger cognate facilitation effects for words with greater orthographic similarity (OS; Dijkstra et al., 2010). Importantly, cognate words do not only differ in terms of OS between languages, but also in the degree of phonological overlap across languages. Recent models, such as the bilingual spelling in alphabetic systems (BAST) model (Tainturier, 2019), propose that the strength of co-activation is mediated by the degree of orthographic and phonological similarity (PS) between the two languages. However, the combined contributions of OS and PS have received little attention.

Most studies focusing on the interplay between OS and PS have been conducted using reading paradigms using strings of letters on the screen (Schwartz et al., 2007; Comesaña et al.,

2012). The fact that the presented input is orthographic can undermine the possible role of phonology on language processing. According to cognitive models of reading (e.g., the dual-route model of reading; Coltheart et al., 2001), a visual stimulus may be decoded through the orthography to phonology conversion (OPC) system where a mapping between graphemes and phonemes occurs (letter-sound correspondence rules). Thus, during silent reading, phonology is activated, but its activation is delayed with respect to the first orthographic analysis. As such, in these kind of reading tasks, processing may be biased toward orthographic decoding. Conversely, writing production paradigms, and especially the writing to dictation task, can provide a useful tool to study the role of phonology and its interplay with orthography. In a writing to dictation task, the first input is phonological [phonology to orthography conversion (POC) system], due to words that are presented by auditory modality (e.g., the dual-route of spelling; Houghton and Zorzi, 2003) and therefore, orthographic activation occurs later than phonological activation (see **Figure 1**).

An effective approach to study the interplay of OS and PS could be the orthogonal manipulation of both variables. Comesaña et al. (2012) divided the cognate condition into four experimental conditions depending on the degree of orthographic and phonological similarity: O + P+ (*bomba*-BOMB), O + P− (*cometa*-COMET), O − P+ (*dança*-DANCE), and O − P− (*laço*-LACE), where the sign “+” indicates high overlap between languages, and the sign “−” indicates low overlap. Twenty-four Portuguese-English bilinguals performed a silent reading task including cognate and non-cognate words during a masked priming paradigm. Participants had to press the space bar to proceed to the next word (i.e., a self-paced reading task). Overall, performance (reaction times) was better for non-cognates than for cognates. Phonological effects were also present but they depended on the degree of orthographic similarity. Thus, cognates with high PS were read faster than cognates with low PS, but these differences were restricted to the high OS conditions (O + P+ vs. O + P−). For low OS cognates, the effect of phonology disappeared. In another study, Schwartz et al. (2007) asked English-Spanish bilinguals to read aloud cognates and non-cognates in both languages in two counterbalanced blocks. The orthogonal manipulation of orthographic and phonological similarity was also included as: O + P+ (*hospital*-HOSPITAL), O + P− (*genuino*-GENUINE), O − P+ (*noCIÓN*-NOTION), and O − P− (*músculo*-MUSCLE). Reading latencies were slower for cognates relative to non-cognates, suggesting an interference effect (from the onset of stimulus presentation to the onset of articulation). In addition, cognate words with high orthographic and phonological similarity (O + P+) were named faster than cognates with high orthographic similarity but low phonological overlap (O + P−). However, there was no difference between O − P+ and O − P−. That is, when the OS between languages was low, there was no PS effect (faster responses for high PS cognates than for low PS cognates). Therefore, the co-activation of phonology seems to be OS-dependent (orthographic autonomy hypothesis; Rapp and Caramazza, 1997). Only when the OS between languages was high was the phonology activated. Importantly, this pattern



of results was observed both in the L2 (Spanish block) and L1 (English block). Hence, cross-language influences were evident during reading in the weaker L2 but also in the stronger L1.

The goal of the current study is to investigate the role of cognate status in bilingual writing production using a writing to dictation task in which a phonological analysis is mandatory. Specifically, we (1) compared performance (reaction time and accuracy) for cognate and non-cognate words in a typing paradigm and (2) examined the effect of orthographic and phonological co-activation in writing performance. To our knowledge, this study is the first to test the effect of orthographic and phonological activation across languages during a writing to dictation task. The critical materials included in this experiment consisted of cognate and non-cognate words (extracted from Schwartz et al., 2007). We included also the orthogonal manipulation of OS and PS: O+P+; O+P−; O−P+; and O−P−. Following previous studies investigating bilingual word recognition, we expected that the cognate facilitation effect (e.g., Costa et al., 2005; Hoshino and Kroll, 2008; Lemhöfer et al., 2008; Dijkstra et al., 2010) would be modulated by orthography, and more importantly also by the phonological overlap across languages. As in Schwartz et al. (2007), we expected that O+P+ would be typed faster than O+P− cognates, as evidence that phonological information is processed.

However, in contrast to previous results, we also expected differences when the orthographic forms of cognates were different (O−P+ vs. O−P−), due to the differences between experimental tasks (see Figure 1). Different from reading studies in which the phonology only has an effect in high OS conditions, in our writing to dictation task, we predicted that the phonology would have an effect for high OS as well as for low OS conditions (significant differences between P+ and P−). In writing to dictation, the first input is phonological, so the phonological processing precedes orthographic processing, and therefore, the phonology would have a direct impact on performance. In this case, the phonological processing would be relatively independent of the orthographic overlap.

In addition to variations in the type of task, phonological and orthographic co-activation may also be dependent on the previous language experience of the bilingual participants. Previous studies have shown that the relationship between L1 and L2 is influenced by L2 competence and by the language learning background (Kroll et al., 2006; Dijkstra et al., 2010). Language experience is characterized by high variability on a range of factors related to language exposure and use (Green and Abutalebi, 2015; Anderson et al., 2018). The nature of the input received during learning has important consequences on language processing (Kroll et al., 2018; Fricke et al., 2019) and language outcomes (Place and Hoff, 2011; Byers-Heinlein,

2013). The quantity, and, even more important, the quality of the input are strong predictors of the language development in bilinguals (Gathercole and Thomas, 2009). In this context, it is fundamental to consider differences between naturalistic and classroom settings (Rothman and Guijarro-Fuentes, 2010). It is well known that L2 learners in a classroom setting receive considerably less oral input than in a naturalistic setting (and of course than native speakers). Qualitative differences in input during learning might serve to explain some asymmetries between L2-learners in classroom and naturalistic environments. The learning background might be especially relevant when examining bilingual writing because writing competence might differ depending on whether L1 or L2 was formally acquired at school, or whether it was learned and used at home where verbal/auditory input exceeds visual/written exposure. These differences could have an important impact on the interplay of orthographic and phonological processing.

In order to address this critical question, we included two groups of English-Spanish bilinguals with different language learning backgrounds: native English speakers who were Spanish learners [late bilinguals (LBs) with formal education in Spanish] and Spanish heritage speakers (HSs) who had acquired English and Spanish at an early age in the household but did not receive a formal education in Spanish. Both groups of participants were immersed in an English dominant context and immersed in English education. The selection of these two groups provides the opportunity for examining the effects of phonological and orthographic co-activation in cognate writing production by English-Spanish bilinguals, who have different background experiences in one of their languages, experience with academic literacy and formal instruction in Spanish and English (LBs) vs. experience with academic literacy and formal instruction just in English (HSs; Carrasco-Ortiz et al., 2019). L1 acquisition is normally characterized by being homogeneous, systematic, and complete. However, the L1 acquisition in the HSs could be unstable and incomplete (Montrul, 2008; Polinsky, 2011). As HSs learn their minority language (L1) at home, and at the same time, they are immersed in a majority language (L2) context (Benmamoun et al., 2013), they receive mainly oral/phonological input during L1-learning (in a naturalistic environment). In contrast, L2 learners are exposed to formal education of reading and writing, but also to oral inputs in an instructed context (e.g., Hytlenstam and Abrahamsson, 2003; Paradis, 2004). Given the higher exposure to oral/phonological input in HSs in comparison with L2 learners, HSs are thought to have a phonological advantage (Chang et al., 2011; Gor, 2014). In addition, studies have also pointed out difficulties in orthographic knowledge in HSs (Elola and Mikulski, 2016) especially during writing tasks (Montrul, 2013). These described differences across bilingual speakers made it possible to expect stronger phonological effects in the HSs than in LBs (faster responses for cognates with high PS than for cognates with low PS), especially during English writing, in which the influence of Spanish phonology is expected. In addition, stronger orthographic effects were expected for LBs relative to HSs, especially during English writing due to their greater familiarity with Spanish orthography. Note that “stronger phonological

effects” mean higher differences between P+ and P– conditions. On the contrary, “stronger orthographic effects” mean higher differences between O+ and O– conditions.

## MATERIALS AND METHODS

### Participants

Forty-eight bilingual students from the University of Florida (United States) participated in the study in exchange for partial course credit. One participant was excluded because he reported central auditory processing disorder. The remaining 47 participants reported normal hearing and normal vision, and they did not report any language or neurological deficits. All participants were able to type using their 10 fingers. They were classified into two experimental groups: 23 LBs and 24 HSs. Both groups were immersed in an English dominant context and they had been educated in the United States.

As data analysis was implemented as mixed-effect regression analysis, we checked if our observations were enough for this type of analysis. Brysbaert and Stevens (2018) recommend “at least 1,600-word observations per condition (e.g., 40 participants, 40 stimuli).” In the current study, observations from 47 participants (23 LBs and 24 HSs) and from 208 words (104 cognates vs. 104 non-cognates) were included. This resulted in 2392 observations for the LBs, and 2,496 observations for the HSs in each condition. However, some of these observations were excluded from analysis due to the data trimming performed to eliminate outliers (see “Results”). Despite this, we had enough observations, with 2,104 observations remaining in the LBs (and 2,170 for non-cognates), and 2,242 observations in the HSs (and 2,238 for non-cognates). This estimation is similar to the ones reported previous studies (Schwartz et al., 2007; Comesaña et al., 2012).

To determine their language dominance and background experiences (experience with academic literacy and formal instruction), all participants completed the language experience and proficiency questionnaire (LEAP-Q; Marian et al., 2007) for both languages, Spanish and English. **Table 1** summarizes the language use and exposure data and the proficiency level of the participants.

The LEAP-Q data show that the LBs were exposed to English earlier than the HSs [age of first exposure (AoA),  $t(45) = -4.541$ ,  $p < 0.001$ ] because they were born into an English-speaking country/family and context. In addition, LBs spent more years in an English-speaking country,  $t(45) = 2.016$ ,  $p = 0.049$ , and LBs spent more years living in a familiar English environment,  $t(45) = 2.177$ ,  $p = 0.035$  than HSs. Importantly, the difference in years of exposure to school context in English was not significant,  $t(45) = 1.716$ ,  $p = 0.093$ . Importantly, the difference in the self-assessed English skills was not significant (speaking, understanding, and reading; all  $ps > 0.05$ ). In order to explore the instructed context of English learning, we analyzed the specific item *reading contribution to learning* (see the question 4 in the LEAP-Q questionnaire: “Please mark how much the following factors contributed to you English/Spanish learning”). The participants rated this item on a scale of 1 to 10. Low

**TABLE 1 |** Mean scores (with standard deviation in parenthesis) for English and Spanish language experience in the LBs and HSs.

| Language version                          |               | English (L1/majority language) |              | Spanish (L2/minority language) |              |
|-------------------------------------------|---------------|--------------------------------|--------------|--------------------------------|--------------|
|                                           |               | LBs (N = 23)                   | HSs (N = 24) | LBs (N = 23)                   | HSs (N = 24) |
| <i>LEAP-Q items</i>                       |               |                                |              |                                |              |
| AoA                                       |               | 0.74 (0.91)                    | 2.71 (1.87)  | 10.69 (3.61)                   | 0.92 (1.32)  |
| Years of exposure                         | Country       | 19.91 (1.16)                   | 18.71 (2.62) | 0.13 (0.62)                    | 6.08 (8.35)  |
|                                           | Family        | 19.65 (1.99)                   | 16.54 (6.57) | 0.87 (2.41)                    | 19.45 (1.59) |
|                                           | School        | 17.95 (2.94)                   | 16.50 (2.87) | 4.56 (5.01)                    | 3.33 (5.94)  |
| Self-assessed capacity (from 1 to 10)     | to speak      | 9.69 (0.55)                    | 9.54 (0.77)  | 6.30 (1.22)                    | 7.08 (1.21)  |
|                                           | to understand | 9.60 (0.78)                    | 9.71 (0.55)  | 7.35 (1.26)                    | 8.25 (1.22)  |
|                                           | to read       | 9.65 (0.57)                    | 9.66 (0.63)  | 7.26 (1.54)                    | 7.08 (1.47)  |
| Reading contribution to learning          |               | 8.89 (1.42)                    | 8.71 (1.49)  | 7.35 (2.27)                    | 5.87 (2.69)  |
| <b>Spanish Writing and Spelling tests</b> |               |                                |              |                                |              |
| PROESC                                    |               |                                |              | 22.43 (1.87)                   | 20.83 (2.91) |
| TECLE                                     |               |                                |              | 35.74 (7.06)                   | 31.96 (5.20) |

AoA, age of acquisition; LBs, late bilinguals; and HSs, Heritage speakers \* $p < 0.05$ ; \*\* $p < 0.01$ .

scores indicate that reading has contributed little to their learning. This score reflects the degree of formal language education in one language, which is thought to be an important requirement for correct learning of orthography (Iniesta et al., 2021). The differences between groups were not significant;  $t(45) = -0.030$ ,  $p = 0.976$ .

Conversely, HSs were exposed earlier to Spanish (AoA) than LBs [ $t(45) = 8.467$ ,  $p < 0.001$ ] because they were born into a Spanish-speaking family. In addition, HSs lived longer than LBs in a Spanish-speaking country [ $t(45) = -3.408$ ,  $p = 0.001$ ] and familiar Spanish environment [ $t(45) = -31.287$ ,  $p < 0.001$ ]. Overall, HSs presented greater exposure to Spanish. However, the difference with respect to years of exposure to school context in Spanish was not significant [ $t(45) = 0.767$ ,  $p = 0.447$ ]. The difference in the self-assessed Spanish skills was significant for speaking [ $t(45) = -2.193$ ,  $p = 0.034$ ] and understanding [ $t(45) = -2.484$ ,  $p = 0.017$ ]. The HSs scored higher on these scales, as expected. However, in the skill more related to formal use of language, reading [ $t(45) = 0.404$ ,  $p = 0.688$ ], there were no differences between groups. As for English, we explored the *reading contribution to learning* for Spanish revealing that HSs had a significantly lower score [ $t(45) = 2.024$ ,  $p = 0.048$ ].

In addition to the self-rated questionnaire, participants also completed a formal standardized test in Spanish for writing and spelling (PROESC – Batería de Evaluación de Los Procesos de Escritura; Cuetos et al., 2002). As part of PROESC, participants completed the ruled-orthography subtest consisting of a pen and paper writing to dictation task of 25 words that included a Spanish spelling rule (Chacón, 1997). For example, in Spanish, all verbs that end in -aba (i.e., cantaba) are spelled with “b” instead of “v.” In addition, all words that end in -aje (chantaje) are spelled with “j” instead of “g.” In addition, participants completed a silent efficiency reading test (TECLE – Test de Eficiencia Lectora; Marín and Carrillo, 1999) including an orthographic decision subtest in which there were sentences with one word missing. Participants had to select the correct word, among 4 options that included semantic, spelling, and phonological distractors, which included subtle letter changes.

In 3 min, the participant had to solve the maximum number of sentences as possible among a total of 64 sentences. A good knowledge of spelling is necessary to select the correct option. The results showed better accuracy in word writing in PROESC for the LBs (mean = 22.43;  $SD = 1.87$ ) than the HSs (mean = 20.83;  $SD = 2.91$ ) out of 25 words in total,  $t(45) = 2.228$ ,  $p = 0.031$ . Additionally, the LBs were more accurate in the TECLE than the HSs (LBs: mean = 35.74;  $SD = 7.06$ ; HSs: mean = 31.96;  $SD = 5.20$ );  $t(45) = 2.095$ ,  $p = 0.042$ .

These results confirmed that, despite the higher speaking and understanding abilities that HSs reported for Spanish in the self-reported questionnaire, no differences in reading skills were evidenced (the fact that in HSs the superiority in speaking and understanding was not extended to reading could indicate the lower skills with the formal aspect of Spanish). Additionally, the LBs had higher orthographic knowledge of Spanish than the HSs in formal standardized test. This provides support to the assumption that HSs might be biased toward phonology and that they might have more difficulties with the more formal aspects of Spanish (including orthographic rules), due to their informal learning background.

## Materials

A total of 208 words in English and their Spanish translations were selected (extracted from Schwartz et al., 2007). English and Spanish items were presented in two independent blocks. Each language block (Spanish or English) was comprised of 104 cognates and 104 non-cognates. Schwartz et al. (2007) classified them according to the OS score (Van Orden et al., 1988; Yates et al., 2003). If the OS was higher than 0.3, this word was classified as cognate. The conditions were matched in logarithmic lexical frequency and the number of letters (Guasch et al., 2013), age of acquisition (AoA; Kuperman et al., 2012; Alonso et al., 2015), concreteness (Duchon et al., 2013; Brysbaert et al., 2014), and orthographic and phonological neighbors (Marian et al., 2012). English-Spanish cognates and non-cognates were presented aurally. The experimental material

was read by a female Puerto Rican Spanish-English bilingual. The material was recorded using a Shure SM57 microphone on a Marantz Solid State Recorder PMD670 (Valdés Kroff et al., 2019). The recorded items were then isolated using a script implemented in PRAAT software (version 5.3.16; Boersma and Weenink, 2012) employing *TextGrids* for segmentation and labeling. In addition, the script added 50 ms of silence at the beginning and 500 ms at the end of each word by default, and it resampled the words so that they were at 44.1 kHz in monoaural. It also rescaled and equated the loudness of the files. **Table 2** shows descriptive statistics for the experimental material.

As in Schwartz et al. (2007) the cognates condition also included the orthogonal manipulation of OS and PS including high (+) and low (−) similarity: O + P+ ( $n=28$ ); O + P− ( $n=31$ ); O − P+ ( $n=19$ ); and O − P− ( $n=26$ ). If the OS was greater than 0.70, the cognate word was classified as high similarity condition. Otherwise, it was classified as low similarity. The PS was calculated subjectively using the following procedure. Pairs of cognate words were auditorily presented to the participants (English monolinguals). The pairs were recorded and spoken by two fluent bilinguals with each member of the pair spoken by a different bilingual. Participants ( $n=29$ ) rated the phonological similarity of cognate pairs on a Likert scale from 1 (no similarity) to 7 (very similar). If the PS was greater than 4, the cognate word was classified as high similarity. Otherwise, it was classified as low similarity (we report norming that were conducted and reported by Schwartz et al., 2007). **Table 3** shows the OS and PS for each condition. Also see

**Table 2** for the information about frequency, number of letters, age of acquisition, concreteness and neighbors relative to these four experimental conditions.

## Procedure

After signing the consent form, participants in both groups performed the writing to dictation task in two independent blocks (Spanish and English). The order of presentation was counterbalanced between participants. The items were randomized (the four conditions of cognates and the condition of non-cognates). Each block began with eight practice trials, followed by the experimental block, with 208 trials in each language. We included a break in the middle of each block, with a duration adaptable to the needs of the participant. The writing to dictation task was conducted on a computer using E-prime version 3.0 (Psychology Software Tools, Pittsburgh, PA). Participants wore headphones to listen to the stimuli and used a standard QWERTY keyboard to type words. Each trial (see **Figure 2**) started with a fixation point (1) which remained on the screen until the auditory stimulus was presented. As soon as the audio terminated, a position bar (2) appeared on the screen indicating that the participants could start to write. Typing was not enabled until the appearance of this position bar. Participants were instructed to type as quickly and accurately as possible. The responses appeared on the screen at the same time as participants were writing.

Importantly, language co-activation in cognate words could be evidenced as facilitation or interference depending

**TABLE 2 |** Characteristics of the experimental stimuli (mean scores with standard deviations in parenthesis).

| Within-language variables |                                  |                                  |                                  |                                  |                                  |                                  |                                  |
|---------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
|                           | Frequency                        | Letters                          | Concreteness                     | AoA                              | PN                               | ON                               | Audio                            |
| <b>English block</b>      |                                  |                                  |                                  |                                  |                                  |                                  |                                  |
| Non-cognates              | 1.41 (0.47)                      | 6.09 (1.26)                      | 3.92 (1.10)                      | 6.53 (2.01)                      | 6.93 (6.99)                      | 4.01 (5.21)                      | 647 (144)                        |
| Cognates                  | 1.42 (0.48)                      | 5.94 (1.41)                      | 3.69 (1.01)                      | 6.95 (2.25)                      | 5.48 (7.59)                      | 3.50 (4.11)                      | 667 (124)                        |
| Statistics                | $t(206)=0.188$ ,<br>$p=0.851$    | $t(206)=-0.837$ ,<br>$p=0.404$   | $t(206)=-1.531$ ,<br>$p=0.127$   | $t(206)=1.43$ ,<br>$p=0.154$     | $t(206)=-1.42$ ,<br>$p=0.157$    | $t(206)=-0.779$ ,<br>$p=0.437$   | $t(206)=1.08$ ,<br>$p=0.203$     |
| O + P+                    | 1.50 (0.46)                      | 6.11 (1.51)                      | 3.91 (0.91)                      | 6.98 (1.47)                      | 5.75 (7.68)                      | 3.35 (3.95)                      | 700 (150)                        |
| O + P−                    | 1.35 (0.48)                      | 5.87 (1.49)                      | 3.53 (0.97)                      | 7.02 (2.67)                      | 5.13 (8.62)                      | 4.03 (4.45)                      | 632 (89)                         |
| O − P+                    | 1.43 (0.34)                      | 5.89 (0.87)                      | 3.80 (1.24)                      | 7.24 (2.04)                      | 4.53 (5.25)                      | 3.36 (2.26)                      | 698 (112)                        |
| O − P−                    | 1.41 (0.61)                      | 5.88 (1.58)                      | 3.56 (0.93)                      | 6.62 (2.61)                      | 6.30 (6.93)                      | 3.84 (4.51)                      | 652 (127)                        |
| Statistics                | $F(3, 104)=0.433$ ,<br>$p=0.729$ | $F(3, 104)=0.171$ ,<br>$p=0.916$ | $F(3, 104)=0.910$ ,<br>$p=0.439$ | $F(3, 104)=0.299$ ,<br>$p=0.826$ | $F(3, 104)=0.232$ ,<br>$p=0.874$ | $F(3, 104)=0.722$ ,<br>$p=0.541$ | $F(3, 104)=1.897$ ,<br>$p=0.204$ |
| <b>Spanish block</b>      |                                  |                                  |                                  |                                  |                                  |                                  |                                  |
| Non-cognates              | 1.37 (0.47)                      | 6.65 (1.72)                      | 4.98 (1.42)                      | 5.28 (2.35)                      | 3.15 (4.45)                      | 2.89 (3.42)                      | 663 (142)                        |
| Cognates                  | 1.37 (0.54)                      | 6.27 (1.53)                      | 4.74 (0.98)                      | 5.85 (3.06)                      | 3.10 (3.88)                      | 2.66 (3.26)                      | 673 (128)                        |
| Statistics                | $t(206)=-1.211$ ,<br>$p=0.227$   | $t(206)=-1.648$ ,<br>$p=0.101$   | $t(206)=-1.21$ ,<br>$p=0.232$    | $t(206)=0.702$ ,<br>$p=0.484$    | $t(206)=-0.074$ ,<br>$p=0.941$   | $t(206)=-0.492$ ,<br>$p=0.623$   | $t(206)=0.516$ ,<br>$p=0.607$    |
| O + P+                    | 1.48 (0.42)                      | 6.35 (1.59)                      | 4.87 (0.91)                      | 5.90 (2.16)                      | 2.96 (3.13)                      | 2.71 (2.59)                      | 698 (146)                        |
| O + P−                    | 1.39 (0.64)                      | 6.16 (1.55)                      | 4.73 (1.14)                      | 5.40 (3.70)                      | 2.53 (3.23)                      | 2.53 (2.78)                      | 657 (119)                        |
| O − P+                    | 1.27 (0.55)                      | 6.21 (1.03)                      | 4.73 (1.11)                      | 6.43 (2.55)                      | 2.94 (3.09)                      | 1.52 (1.64)                      | 651 (111)                        |
| O − P−                    | 1.42 (0.48)                      | 6.58 (2.13)                      | 4.62 (0.89)                      | 5.56 (2.86)                      | 3.11 (3.88)                      | 2.61 (3.23)                      | 680 (132)                        |
| Statistics                | $F(3, 104)=0.573$ ,<br>$p=0.634$ | $F(3, 104)=0.335$ ,<br>$p=0.800$ | $F(3, 104)=0.227$ ,<br>$p=0.877$ | $F(3, 104)=0.550$ ,<br>$p=0.649$ | $F(3, 104)=0.757$ ,<br>$p=0.521$ | $F(3, 104)=1.496$ ,<br>$p=0.220$ | $F(3, 104)=0.747$ ,<br>$p=0.526$ |

AoA, age of acquisition; PN, phonological neighbors; ON, orthographic neighbors; Audio, audio duration; +, high similarity; and −, low similarity

**TABLE 3 |** Orthographic and phonological similarity across experimental conditions (mean scores with standard deviations in parenthesis).

| Cross-linguistic variables |                                      |                                      |
|----------------------------|--------------------------------------|--------------------------------------|
|                            | OS                                   | PS                                   |
| Non-cognates               | 0.14 (0.09)                          | –                                    |
| Cognates                   | 0.74 (0.23)                          | –                                    |
| Statistics                 | $t(206) = 24.21$ ,<br>$p = 0.000$    |                                      |
| (1) O + P +                | 0.92 (0.12)                          | 5.31 (0.91)                          |
| (2) O + P –                | 0.88 (0.13)                          | 2.84 (0.67)                          |
| (3) O – P +                | 0.50 (0.13)                          | 5.02 (0.71)                          |
| (4) O – P –                | 0.54 (0.17)                          | 2.84 (0.73)                          |
| Statistics                 | $F(3, 104) = 60.45$ ,<br>$p = 0.000$ | $F(3, 104) = 81.34$ ,<br>$p = 0.000$ |
| Post-hoc comparisons       | 1 vs. 2                              | $t(57) = 0.931$ ,<br>$p = 0.356$     |
|                            | 1 vs. 3                              | $t(45) = 11.11$ ,<br>$p = 0.000$     |
|                            | 1 vs. 4                              | $t(52) = 9.27$ ,<br>$p = 0.000$      |
|                            | 2 vs. 3                              | $t(48) = 9.84$ ,<br>$p = 0.000$      |
|                            | 2 vs. 4                              | $t(55) = 8.406$ ,<br>$p = 0.000$     |
|                            | 3 vs. 4                              | $t(43) = -0.963$ ,<br>$p = 0.341$    |
|                            |                                      | $t(57) = 11.87$ ,<br>$p = 0.000$     |
|                            |                                      | $t(45) = 1.209$ ,<br>$p = 0.233$     |

OS, orthographic similarity; PS, phonological similarity. We report norming that were conducted and reported by Schwartz et al. (2007). Data from non-cognates words were not available in the original research.

on whether co-activation occurs at a lexical or sublexical level (Muscalu and Smiley, 2018; Iniesta et al., 2021), or depending on whether co-activation occurs in a more initial and central process (lexical retrieval), or in a more posterior or peripheral process (Purcell et al., 2011). For this reason, the reaction time (RT) and accuracy (ACC) of the typing response were collected in two different temporal moments associated with lexical and sublexical processing (see Muscalu and Smiley, 2018 and Iniesta et al., 2021 for a similar procedure): from the offset of the stimulus to the first keypress (first key performance) (3) and from the first keypress to the press of the space bar key (rest of the word performance) (4). These two measures have been associated with lexical and sublexical processing, respectively, and therefore allowed us to pinpoint the time course and level of linguistic analysis at which our effects occurred. Considering that the experiment was carried out with an English keyboard, the participants received explicit instructions not to write the diacritical marks during the Spanish block. In addition, one word included a “ñ” grapheme. The participants were instructed to press the key adjacent to the “l,” which would be the natural position of the ñ on a Spanish keyboard. Between trials, there was a black screen for 1000 ms (5).

Between the English and Spanish blocks of the writing to dictation task, participants completed the LEAP-Q questionnaire (Marian et al., 2007) for both languages (Spanish and English), and the two Spanish assessment tests (PROESC and TECLE,

see the participants section, for more information). Overall, the experimental session lasted approximately 60 min. The study was conducted in accordance with the ethical standards approved by the University of Florida Institutional Review Board (IRB): protocol #2019–02427.

## RESULTS

For the writing to dictation task, the ACC and the RTs for correct responses were calculated for each participant and condition for the first keystroke and the rest of the word separately. Response times above or below 2.5 SD from each participant’s mean were eliminated from the analysis [first key performance: 3.31% (English)/4.39% (Spanish) of the items of the LBs and 3.92% (English)/4.49% (Spanish) from the HSs; rest of the word performance: 4.15% (English)/4.87% (Spanish) of the items of the LBs and 4.88% (English)/5.58% (Spanish) from the HSs]. Accuracy was determined based on a strict criterion for correct (1) vs. incorrect (0) scores. Clear typographical errors were also considered as correct (e.g., helicqopter. In this case, the key “q” is not necessary and it is not surrounding any target key). In the same way, errors derived from accentuation in Spanish were also considered correct. Although the instructions explicitly indicated not to type the accent marks, some participants made mistakes trying to type them, and we also considered these words as correctly typed (e.g., m^aquina, the Spanish word for machine). Note that there were only eight observations in this special situation.

Following previous studies, two independent analyses were conducted to explore (Schwartz et al., 2007; Comesaña et al., 2012): (1) the overall effect of language and cognate status in the performance of both groups of participants and (2) the impact of OS and PS in cognates.

A mixed-model analysis was performed using the R lme4 package (R Core Team, 2017; Bates et al., 2015) and including the function with a “Kenward-Roger” modification for  $F$ -tests (Halekoh and Højsgaard, 2014) in order to include the random effects in the analysis (Luke, 2017). The model for the first analysis (overall effect of language and cognate status) was conducted with Group (LBs vs. HSs), Language (English vs. Spanish), and Condition (Cognates vs. non-cognates) as fixed factors and Participants and Items as random effects for first key and rest of the word performances. For the second analysis (the impact of OS and PS), the model included Group (LBs vs. HSs), Language (English vs. Spanish), OS (+ vs. –), and PS (+ vs. –) as fixed factors and Participants and Items as random effects both for first key and rest of the word performances. Participants and Items were included as random intercepts, random slopes were not included because a simplification of the maximal following the convergence of models (Barr et al., 2013). When a two-way interaction was found, a post-hoc  $t$ -test using Tukey’s multiple comparison correction was implemented using the R function lsmeans. When a three-way interaction (or above) was significant, a new model exploring this

specific interaction was performed, also including participants and items as random effects. Finally,  $p$ -values were reported by the *anova* function of the LmerTestR-package. Full models' summary is available from the OSF repository: [https://osf.io/bkhvj/?view\\_only=325b38c094ff41749f2db2a9ef608286](https://osf.io/bkhvj/?view_only=325b38c094ff41749f2db2a9ef608286).

## The Overall Effect of Language and Cognate Status

**Table 4** summarizes the results (RTs and ACC) obtained in the writing to dictation task as a function of Group (LBs vs. HSs), Language (English vs. Spanish), and Condition (cognates vs. non-cognates).

## First Key Performance

### Latency

For *first key* latencies (RTs), the main effect of Language was significant,  $F(1, 478.9) = 76.38, p < 0.001$ . Responding in English (mean = 638 ms) was faster than responding in Spanish (mean = 757 ms). The main effect of Condition was also significant,  $F(1, 349.4) = 3.95, p = 0.047$ . Cognates were responded to slower (mean = 716 ms) than non-cognates (mean = 679 ms).

The interaction between Group and Language was also significant [ $F(1, 17185.1) = 51.54, p < 0.001$ ]. For both groups, the differences between Spanish and English were significant [LBs:  $t(17182.3) = -10.67, SE = 14.6, p < 0.001$ ; HSs:  $t(17186.4) = -5.66, SE = 14.5, p < 0.001$ ], but the magnitude of

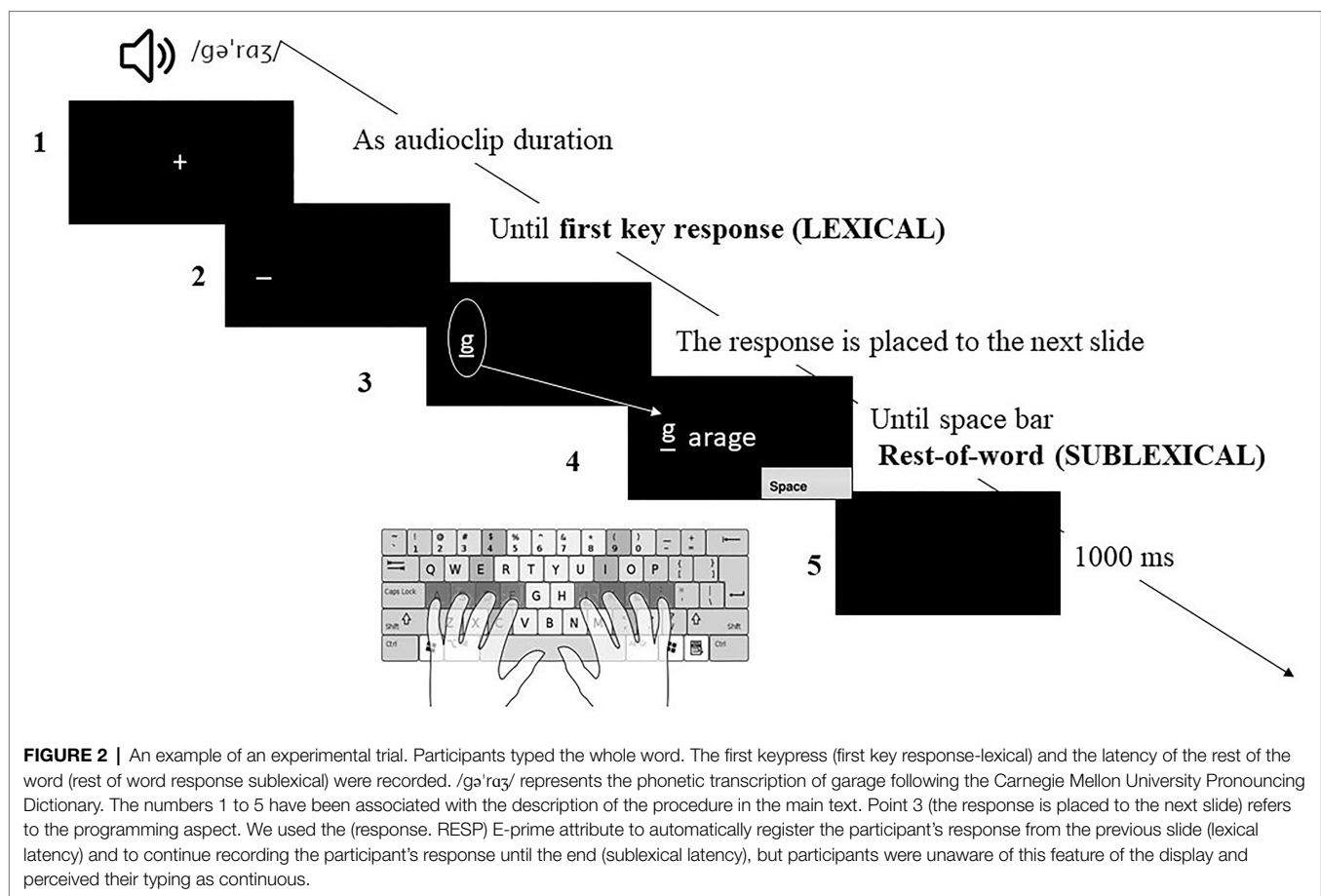


**TABLE 4 |** Mean scores (with standard errors in parenthesis) in the writing to dictation for the overall effect of language and cognate status in each participant group (analysis 1).

|     |              | First key     |               |               |               | Rest of word  |               |               |               |
|-----|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|     |              | English       |               | Spanish       |               | English       |               | Spanish       |               |
|     |              | LBs           | HSs           | LBs           | HSs           | LBs           | HSs           | LBs           | HSs           |
| RTs | Cognates     | 643 (29.8)    | 713 (29.3)    | 764 (29.8)    | 744 (29.3)    | 1182 (58.1)   | 1325 (57.3)   | 1318 (58.0)   | 1550 (57.3)   |
|     | Non-cognates | 562 (31.3)    | 635 (30.9)    | 753 (31.4)    | 768 (31.0)    | 1085 (63.2)   | 1254 (62.4)   | 1549 (63.3)   | 1852 (62.6)   |
| ACC | Cognates     | 0.944 (0.009) | 0.946 (0.009) | 0.943 (0.009) | 0.956 (0.008) | 0.832 (0.019) | 0.817 (0.018) | 0.868 (0.019) | 0.841 (0.018) |
|     | Non-cognates | 0.966 (0.010) | 0.971 (0.010) | 0.924 (0.010) | 0.944         | 0.898 (0.021) | 0.884 (0.021) | 0.829 (0.022) | 0.812 (0.022) |

the differences was greater in the LBs (Spanish: 759–English: 602=157 ms) than in the HSs (Spanish 756 – English: 674=82 ms). The interaction between Language and Condition was also significant [ $F(1, 478.8)=9.93, p=0.002$ ] with cognates being slower (mean=678 ms) than the non-cognates [mean=598 ms;  $t(477.2)=3.48, SE=23.0, p<0.001$ ] in the English block. However, these differences were not significant in the Spanish block [mean of cognates=754 ms; mean of non-cognates=760 ms;  $t(475.2)=-0.26, SE=23.1, p=0.791$ ]. No other main effects or interactions were significant (all  $ps>0.05$ ).

### Accuracy

For *first key* accuracy, there was a main effect of Group,  $F(1, 46.3)=4.25, p=0.045$ , with higher accuracy for HSs (mean=0.954) than for LBs (mean=0.944). The main effect of Language was also significant,  $F(1, 492.7)=4.22, p=0.040$ , such that accuracy in English (mean=0.957) was higher than in Spanish (mean=0.942).

A Group  $\times$  Language interaction was also significant,  $F(1, 18076.1)=5.35, p=0.021$ . For LBs, the difference between English (mean=0.955) and Spanish (mean=0.945) was significant [ $t(18072.3)=2.74, SE=0.008, p=0.006$ ], whereas for HSs, it was not [English mean=0.959; Spanish mean=0.950;  $t(18072.8)=1.09, SE=0.007, p=0.272$ ]. The Language  $\times$  Condition interaction was also significant,  $F(1, 492.7)=6.79, p=0.009$ , showing that for the English block, cognates (mean=0.945) were less accurate than non-cognates (mean=0.969),  $t(491.1)=-1.989, SE=0.012, p=0.046$ . In contrast, for the Spanish block, the difference between cognates and non-cognates was not significant [cognates mean=0.949, non-cognates mean=0.934;  $t(490.8)=1.19, SE=0.013, p=0.232$ ]. No other main effects or interactions were significant (all  $ps>0.05$ ).

### Rest of the Word Performance

#### Latency

Regarding the RTs of the *rest of the word*, there was a main effect of Group,  $F(1, 46.9)=10.79, p=0.002$ . LBs (mean=1284 ms) showed faster responses than the HSs (mean=1495 ms). There was also a main effect of Language,  $F(1, 428.8)=124.66, p<0.001$ . The responses in English (mean=1211 ms) were faster than Spanish (mean=1567 ms). Similarly, the main effect of Condition was significant,  $F(1, 354.2)=4.107, p=0.043$ . Cognates (1344 ms) were typed faster than non-cognates (1435 ms).

The interaction between Group and Language,  $F(1, 15244.1)=38.35, p<0.001$ , was also significant. For both groups, the differences between Spanish and English were significant [LB:  $t(15238.1)=-9.05, SE=33.1, p<0.001$ ; HS:  $t(15245.3)=-12.44, SE=33.1, p<0.001$ ]. However, the magnitude of the difference was greater for the HSs (Spanish 1701 – English: 1289=412 ms) than for the LBs (Spanish: 1434 – English: 1134=300 ms). The interaction between Group  $\times$  Condition was also significant,  $F(1, 15242.3)=7.17, p=0.007$ . Thus, for LBs, there were no differences between cognates (mean=1250 ms) and non-cognates [mean=1357 ms;  $t(15241.6)=-1.46, SE=46.0, p=0.143$ ], whereas these differences were significant in the

HSs [mean of cognates=1437 ms; mean of non-cognates=1553 ms;  $t(15239.4)=-2.51, SE=46, p=0.012$ ]. The interaction between Language and Condition was also significant,  $F(1, 428.6)=30.37, p<0.001$ , such that in the English block, there were no differences between cognates (mean=1254 ms) and non-cognates [mean=1169 ms;  $t(427.3)=1.52, SE=55.2, p=0.128$ ], whereas in the Spanish block, cognates (mean=1434 ms) were faster than non-cognates [mean=1701 ms;  $t(428.4)=-4.83, SE=55.3, p<0.001$ ]. The three-way interaction was not significant [Group  $\times$  Language  $\times$  Condition,  $F(1, 15242.5)=1.61, p=0.204$ ].

### Accuracy

For *rest of the word* accuracy, no main effects were significant; Group,  $F(1, 46.5)=2.25, p=0.139$ ; Language,  $F(1, 443.9)=1.90, p=0.168$ ; and Condition,  $F(1, 363.1)=0.63, p=0.427$ .

However, the Language  $\times$  Condition interaction was significant,  $F(1, 443.9)=11.53, p<0.001$ . In the English block, cognates (mean=0.824) were less accurate than non-cognates (mean=0.891),  $t(442.6)=-2.618, SE=0.025, p=0.008$ , whereas the differences in the Spanish block were not significant [cognates mean=0.854, non-cognates mean=0.821;  $t(443.7)=1.315, SE=0.025, p=0.188$ ]. No other interactions were significant (all  $ps>0.05$ ).

### Summary of the Language and Cognate Status Analysis

The *first key* responses were slower for Spanish (L2) than for English (L1), although this effect was modulated by subtle differences in language experience (e.g., LBs were slower in Spanish than in English to a greater extent than the HSs). In addition, LBs were more accurate in English than in Spanish, but these language differences in accuracy were not present in HSs). In addition, both groups showed similar patterns of cognate effects, with cognate interference being evident in English (L1), but absent in Spanish (L2), in latency, and accuracy. For the *rest of word*, response times differed for language, group, and condition: Responses were slower for Spanish (L2) than for English (L1), although this effect was modulated by the differences in language experience (e.g., HSs were slower in Spanish than in English to a greater extent than the LBs). Writing cognate words were faster than writing non-cognate words, but this facilitatory effect showed some nuanced relations with language (only present in Spanish when looking at response times). Importantly, the group-by-condition interaction indicated that the facilitatory effect was only present for the HSs. However, in English (L1), writing cognates were less accurate than writing non-cognate words in both groups, revealing a similar cognate interference effect to that found for the first key.

### The Impact of Orthographic (OS) and Phonological Similarity (PS) in Cognates

Figure 3 (for latency) and Figure 4 (for accuracy) summarize the results obtained in the writing to dictation task in relation to a new analysis including four factors: Group (LBs vs. HSs), Language (English vs. Spanish), OS (High vs. Low), and PS

(High vs. Low) within the cognate condition. In addition, a summary of statistics has been included in **Table 5** (main effects and interactions). In the following subsections, we further analyze the significant effects reported in **Table 5**.

## First Key Performance

### Latency

Regarding the latency (RTs) of the *first key*, the main effect of Language was significant. The responses in English (mean=658ms) were faster than in Spanish (mean=763ms). The main effect of PS was also significant. Cognates with high PS (mean=648ms) were typed faster than of cognates with low PS (mean=773 ms).

The interaction between Group and Language was significant. The differences between Spanish and English were significant in both groups [LBs:  $t(8612.6) = -9.23$ ,  $SE = 16.1$ ,  $p < 0.001$ ; HSs:  $t(8610.3) = -3.86$ ,  $SE = 15.9$ ,  $p < 0.001$ ], but the magnitude

of the differences was greater in the LBs (Spanish: 772 – English: 623=149ms) than in the HSs (Spanish 754 – English: 692=22ms). The interaction between Language and OS was also significant, indicating that in the English block, there were no significant differences between cognates with high OS (mean=659ms) and cognates with low OS [mean=657ms;  $t(1680.9) = 0.07$ ,  $SE = 30.7$ ,  $p = 0.945$ ], whereas in the Spanish block, these differences were significant (mean of cognates with high OS=717ms; mean of cognates with low OS=808ms;  $t(1681.7) = -2.97$ ,  $SE = 30.7$ ,  $p = 0.002$ ).

The three-way interaction between Group, Language, and PS was also significant. In order to explore this interaction, we performed a specific model (Language\*PS) for each group separately. Here, we wanted to examine the interaction between Language and PS separately for the LBs and the HSs in order to examine the PS effect in each language, across the two language background profiles. The analysis performed in the LBs indicated a main effect of Language [ $F(1, 1797.2) = 51.01$ ,

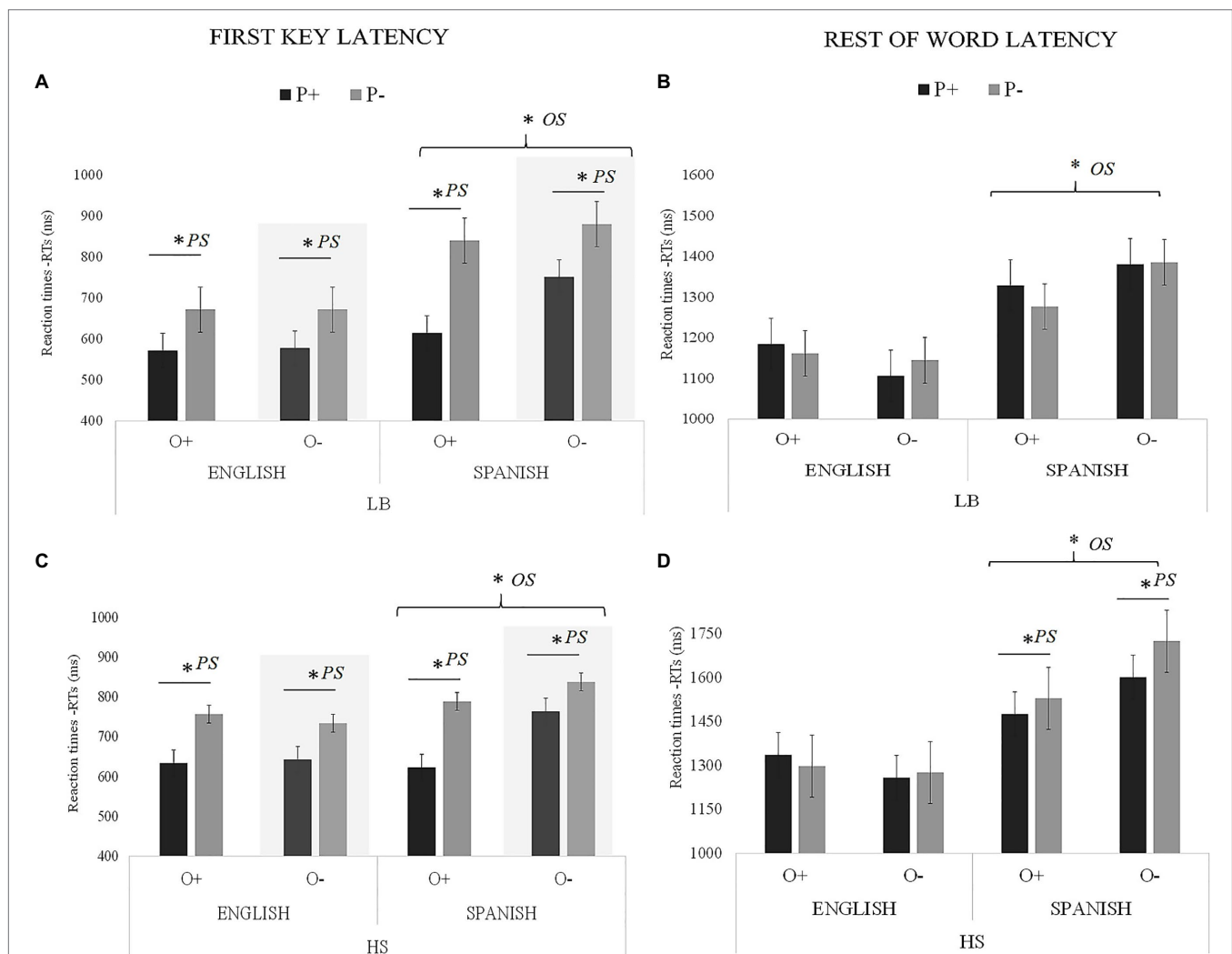
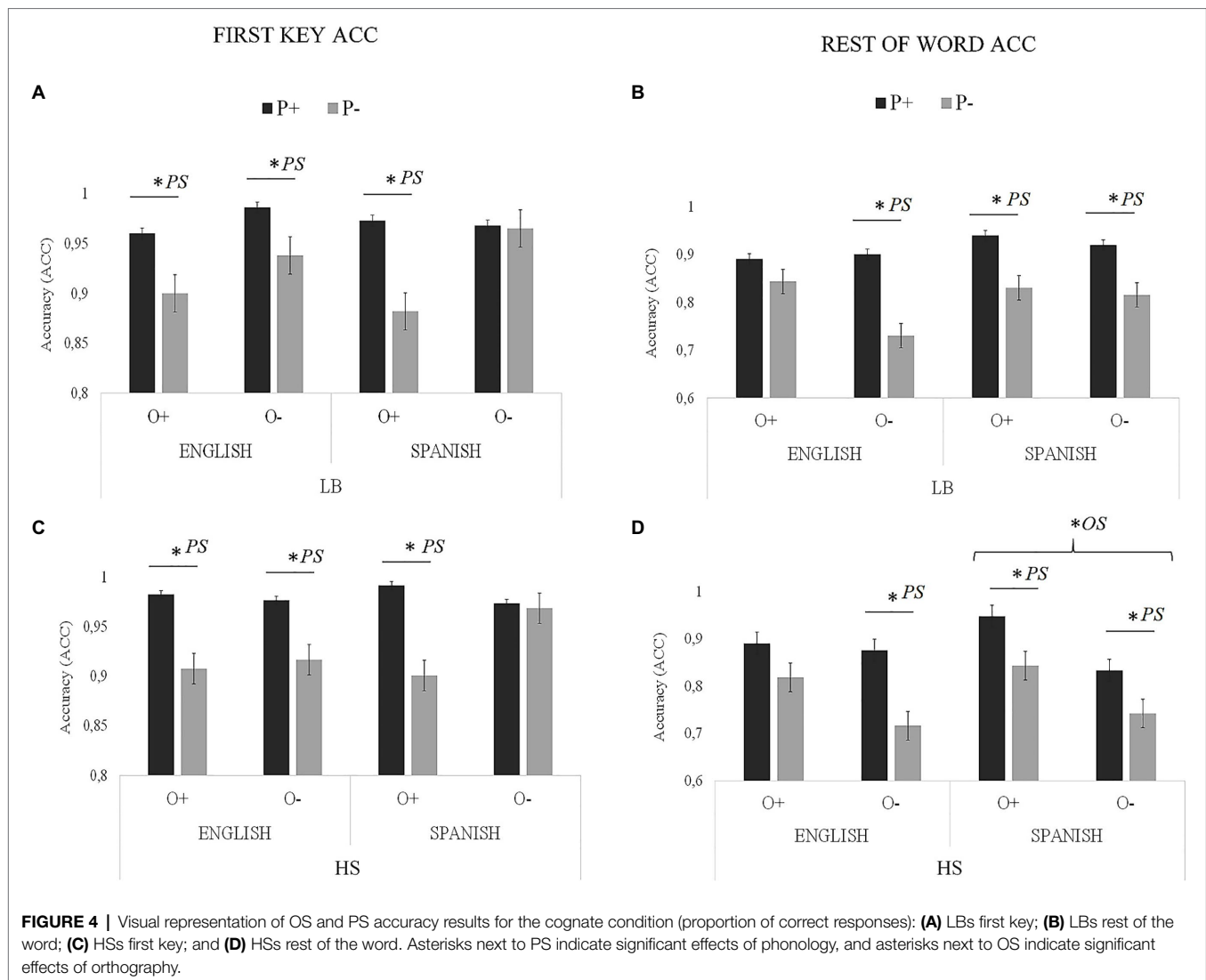


**FIGURE 3 |** Visual representation of OS and PS latency results for the cognate condition (milliseconds): **(A)** LBs first key; **(B)** LBs rest of the word; **(C)** HSs first key; and **(D)** HSs rest of the word. Asterisks next to PS indicate significant effects of phonology, and asterisks next to OS indicate significant effects of orthography.



$p < 0.001$ ] and PS [ $F(1, 149.28) = 20.73, p = 0.003$ ]. In addition, the Language  $\times$  PS interaction was significant,  $F(1, 1796.97) = 9.75, p = 0.002$ . During the English version of the task, cognates with high PS (mean = 591 ms) were typed faster than of cognates with low PS (mean = 672 ms),  $t(1793.45) = -2.85, SE = 34.6, p = 0.043$ . During the Spanish version of the task, cognates with high PS (mean = 662 ms) were also typed faster than of cognates with low PS (mean = 853 ms),  $t(1789.7) = -5.51, SE = 34.7, p < 0.001$ . Although in both languages there were differences between conditions, the magnitude of the differences was greater in Spanish (191 ms) than in English (81 ms). In HSs, there was a main effect of Language,  $F(1, 2839.1) = 9.68, p = 0.002$ . The responses in English (mean = 631 ms) were faster than in Spanish (mean = 757 ms). The main effect of PS was also significant,  $F(1, 148.85) = 18.44, p < 0.001$ . Thus, cognates with high PS (mean = 627 ms) were typed faster than cognates with low PS (mean = 762 ms) but the Language  $\times$  PS interaction was not significant [ $F(1, 2838.04) = 1.49, p = 0.22$ ]. No other interactions were significant.

### Accuracy

Regarding the accuracy (ACC) of the *first key*, the main effect of PS was significant. The accuracy of cognates with high PS (mean = 0.977) was higher than of cognates with low PS (mean = 0.922).

The Group  $\times$  OS interaction was also significant. In the LBs, the difference between cognates with high OS (mean = 0.929) and cognates with low OS (mean = 0.964) was significant [ $t(9076.8) = -2.36, SE = 0.015, p = 0.018$ ], but in the HSs were not significant [O + mean = 0.946; O- mean = 0.959;  $t(9075.4) = -0.882, SE = 0.015, p = 0.377$ ].

The three-way interaction between Language, OS, and PS was also significant. In order to explore the interaction, we performed a specific model (OS\*PS) for each language separately. Here, we wanted to examine the interaction between OS and PS separately for each language in order to examine the interplay of OS and PS effect in each language. In the analysis performed in the English block, there was a main effect of PS,  $F(1, 92.01) = 6.74, p = 0.011$ . The accuracy of

**TABLE 5 |** Summary of results (main effects and interactions) of the OS and PS in the cognate words condition (analysis 2).

| Effects          | First key                                               |                                                        | Rest of the word                                        |                                                         |
|------------------|---------------------------------------------------------|--------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|
|                  | Latency                                                 | ACC                                                    | Latency                                                 | ACC                                                     |
| Group            | $F(1, 47)=0.42, p=0.515$                                | $F(1, 48)=1.44, p=0.235$                               | <b><math>F(1, 46.9)=8.02, p=0.006^*</math></b>          | <b><math>F(1, 47)=4.63, p=0.036^*</math></b>            |
| Lang             | <b><math>F(1, 1,684)=57.02, p&lt;0.001^{**}</math></b>  | $F(1, 1665.2)=0.85, p=0.358$                           | <b><math>F(1, 2,505)=86.21, p&lt;0.001^{**}</math></b>  | <b><math>F(1, 3426.8)=4.57, p=0.032^*</math></b>        |
| OS               | $F(1, 155.3)=2.65, p=0.105$                             | $F(1, 158.3)=2.85, p=0.093$                            | $F(1, 157.2)=0.33, p=0.565$                             | $F(1, 161.3)=3.68, p=0.056$                             |
| PS               | <b><math>F(1, 155.3)=20.91, p&lt;0.001^{**}</math></b>  | <b><math>F(1, 158.3)=14.18, p&lt;0.001^{**}</math></b> | $F(1, 157.2)=0.06, p=0.806$                             | <b><math>F(1, 161.3)=12.22, p&lt;0.001^{**}</math></b>  |
| Group*Lang       | <b><math>F(1, 8613.1)=30.34, p&lt;0.001^{**}</math></b> | $F(1, 9079.3)=1.81, p=0.179$                           | <b><math>F(1, 7663.3)=13.66, p&lt;0.001^{**}</math></b> | $F(1, 9069.8)=2.23, p=0.136$                            |
| Group*OS         | $F(1, 8613.4)=0.01, p=0.898$                            | <b><math>F(1, 9079.7)=6.94, p=0.008^*</math></b>       | $F(1, 7662.1)=2.15, p=0.143$                            | <b><math>F(1, 9069.5)=14.58, p&lt;0.001^{**}</math></b> |
| Group*PS         | $F(1, 8613.6)=2.23, p=0.135$                            | $F(1, 9080.2)=0.70, p=0.403$                           | $F(1, 7662.5)=3.19, p=0.074$                            | $F(1, 9069.5)=0.01, p=0.905$                            |
| Lang*OS          | <b><math>F(1, 1683.6)=11.30, p&lt;0.001^{**}</math></b> | $F(1, 1665.2)=1.04, p=0.309$                           | <b><math>F(1, 2502.8)=10.47, p=0.001^*</math></b>       | $F(1, 3426.6)=0.09, p=0.759$                            |
| Lang*PS          | $F(1, 1683.5)=2.86, p=0.091$                            | $F(1, 1665.1)=0.84, p=0.359$                           | $F(1, 2504.2)=0.42, p=0.516$                            | $F(1, 3426.5)=0.16, p=0.686$                            |
| OS*PS            | $F(1, 155.3)=1.07, p=0.302$                             | $F(1, 158.3)=3.01, p=0.085$                            | $F(1, 157.2)=0.24, p=0.622$                             | $F(1, 161.3)=0.611, p=0.436$                            |
| Group*Lang*OS    | $F(1, 8,613)=0.28, p=0.597$                             | $F(1, 9079.5)=0.95, p=0.329$                           | $F(1, 7662.2)=2.49, p=0.114$                            | <b><math>F(1, 9069.5)=10.56, p=0.001^*</math></b>       |
| Group*Lang*PS    | <b><math>F(1, 8,613)=4.61, p=0.032^*</math></b>         | $F(1, 9,079)=0.49, p=0.485$                            | <b><math>F(1, 7662.4)=6.15, p=0.013^*</math></b>        | $F(1, 9069.9)=0.37, p=0.539$                            |
| Group*OS*PS      | $F(1, 8613.3)=0.09, p=0.755$                            | $F(1, 9079.1)=0.01, p=0.963$                           | $F(1, 7662.4)=0.01, p=0.941$                            | $F(1, 9069.7)=0.71, p=0.399$                            |
| Lang*OS*PS       | $F(1, 1683.6)=1.85, p=0.173$                            | <b><math>F(1, 1665.2)=6.36, p=0.012^*</math></b>       | $F(1, 2504.1)=0.01, p=0.965$                            | <b><math>F(1, 3426.6)=5.58, p=0.018^*</math></b>        |
| Group*Lang*OS*PS | $F(1, 8,613)=0.23, p=0.631$                             | $F(1, 9079.9)=0.01, p=0.936$                           | $F(1, 7662.2)=0.02, p=0.889$                            | $F(1, 9069.7)=0.31, p=0.578$                            |

Lang, language; OS, orthographic similarity; and PS, phonological similarity. \* $p<0.05$ ; \*\* $p<0.01$ . Significant effects are bolded

cognates with high PS (mean=0.984) was higher than of cognates with low PS (mean=0.928). However, the main effect of OS  $F(1, 92.01)=1.10, p=0.296$  and the OS  $\times$  PS interaction  $F(1, 92.01)=0.20, p=0.65$  were not significant. The analysis performed in the Spanish block indicated that there was no main effect of OS,  $F(1, 100.98)=0.1941, p=0.166$ , but the main effect for PS  $F(1, 100.99)=7.84, p=0.006$  and OS  $\times$  PS interaction were significant  $F(1, 100.89)=3.93, p=0.048$ . This interaction indicated that in the high OS condition, there were differences between the P+ (mean=0.977) and P- (mean=0.873) conditions;  $t(100.34)=3.622, SE=0.028, p<0.001$ . However, the difference between P+ (mean=0.965) and P- (mean=0.946) in the low OS condition was not significant;  $t(100.48)=0.573, SE=0.033, p=0.567$ .

## Rest of the Word Performance

### Latency

Regarding the latency (RTs) for the *rest of word*, the main effects of Group and Language were significant, indicating that the responses in the LBs (mean=1246 ms) were faster than in the HSs (mean=1438 ms) and that the responses in English (mean=1221 ms) were faster than those in Spanish (mean=1463 ms).

The interaction between Group and Language was significant. Differences between Spanish and English were significant for both groups [LBs:  $t(7661.8)=-6.630, SE=29.2, p<0.001$ ; HSs:  $t(7662.8)=-9.944, SE=29.2, p<0.001$ ], but the magnitude of the differences was greater in the HSs (Spanish: 1583 – English: 1293=290 ms) than in the LBs (Spanish 1,343 – English: 1150=193 ms). The interaction between Language and OS was also significant, indicating that the cognates with high OS were typed faster than cognates with low OS, but only in Spanish [O+ mean=1403 ms; O- mean=1523 ms;  $t(2489.3)=-2.963, SE=80.1, p=0.038$ ]. In English, there were no differences [O+ mean=1245 ms; O- mean=1195 ms;  $t(2493.5)=1.012, SE=81.4, p=0.472$ ].

The three-way interaction between Group, Language, and PS was also significant. To explore this interaction, we performed a specific model (Language\*PS) for each group separately. Here, we wanted to examine the interaction between language and PS separately for the LBs and the HSs in order to examine the PS effect in each language, across the two language background profiles. The analysis in the LBs indicated that there was a main effect of Language [ $F(1, 3003.7)=46.22, p<0.001$ ], indicating that the responses in English (mean=1,163 ms) were faster than those in Spanish (mean=1,331 ms). The main effect of PS [ $F(1, 152.46)=0.08, p=0.772$ ] and the Language  $\times$  PS interaction [ $F(1, 3003.22)=0.02, p=0.874$ ] were not significant. The analysis for the HSs showed a main effect of Language [ $F(1, 3003.7)=66.17, p<0.001$ ], indicating that the responses in English (mean=1316 ms) were faster than those in Spanish (mean=1553 ms). However, the main effect of PS was not significant [ $F(1, 155.43)=0.36, p=0.545$ ]. The Language  $\times$  PS interaction was significant [ $F(1, 3361.422)=3.74, p=0.039$ ], so that in the English block, the difference between cognates with high PS (mean=1314 ms) and cognates with low PS (mean=1319 ms) was not significant,  $t(3002.3)=-0.069, SE=78.2, p=0.999$ , whereas in the Spanish block, the difference between cognates with high PS (mean=1512 ms) and cognates with low PS (mean=1615 ms) was significant,  $t(7664.1)=-2.55, SE=77.9, p=0.019$ .

### Accuracy

Regarding the ACC of the *rest of word*, the main effect of Group was significant, with higher accuracy for the LBs (mean=0.858) than for the HSs (mean=0.833). The main effect of Language was also significant, indicating higher ACC in Spanish (mean=0.858) than in English (0.833). The main effect of PS was also significant. The accuracy of cognates with high PS (mean=0.899) was higher than of cognates with low PS (mean=0.792).

The Group  $\times$  OS interaction was significant, so that in the LBs, the difference between cognates with high OS (mean=0.875) and cognates with low OS (mean=0.841) was not significant [ $t(9067.7)=1.097$ ,  $SE=0.034$ ,  $p=0.273$ ], whereas this difference was significant for the HSs [O+ mean=0.874; O– mean=0.791;  $t(9068.5)=2.661$ ,  $SE=0.031$ ,  $p=0.008$ ].

The three-way interaction between Group, Language, and OS was significant. To explore this interaction, we performed a specific model (Language  $\times$  OS) for each group separately. The analysis in the LBs indicated that there was a main effect of Language, [ $F(1, 1664.11)=5.05$ ,  $p=0.024$ ], indicating higher ACC in Spanish (mean=0.870), than in English (0.835). The main effect of OS [ $F(1, 160.24)=2.21$ ,  $p=0.138$ ] and the Language  $\times$  OS interaction were not significant [ $F(1, 1664.08)=1.28$ ,  $p=0.257$ ]. The analysis for the HSs showed that the main effect of Language was not significant [ $F(1, 1740.22)=0.08$ ,  $p=0.768$ ]; however, the main effect of OS [ $F(1, 161.08)=6.81$ ,  $p=0.009$ ] and Language  $\times$  OS interaction were significant [ $F(1, 1740.08)=3.77$ ,  $p=0.05$ ]. The interaction indicated that in the English block, the difference between cognates with high OS (mean=0.851) and cognates with low OS (mean=0.796) was not significant; [ $t(1738.7)=1.499$ ,  $SE=0.037$ ,  $p=0.438$ ], whereas in the Spanish block, this difference was significant [O+ mean=0.887; O– mean=0.769;  $t(1736.5)=3.198$ ,  $SE=0.037$ ,  $p=0.008$ ].

The three-way interaction between Language, OS, and PS was significant. We explored this interaction by a specific model (OS  $\times$  PS) for each language separately. The analysis in the English block showed no main effect of OS,  $F(1, 103.69)=0.62$ ,  $p=0.431$ , but the main effect of PS  $F(1, 103.69)=10.84$ ,  $p=0.001$  and the OS  $\times$  PS interaction were significant,  $F(1, 103.69)=4.32$ ,  $p=0.023$ . This interaction indicated that for the high OS condition, there were no differences between the high PS (mean=0.919) and the low PS (mean=0.840) conditions;  $t(102.6)=1.375$ ,  $SE=0.057$ ,  $p=0.515$ . However, there were differences between high PS (mean=0.917) and low PS (mean=0.740) in the low OS condition;  $t(101.7)=3.163$ ,  $SE=0.066$ ,  $p=0.008$ . The analysis in the Spanish block showed main effects of OS, [ $F(1, 102.85)=6.20$ ,  $p=0.014$ ], with higher accuracy for cognates with high OS (mean=0.868) than for cognates with low OS (mean=0.799). The main effect of PS was also significant,  $F(1, 102.85)=6.78$ ,  $p=0.012$ , so that accuracy in cognates with high PS (mean=0.871) was higher than in cognates with low PS (mean=0.777). The OS  $\times$  PS interaction was not significant,  $F(1, 102.85)=0.336$ ,  $p=0.563$ . No other effects or interactions reached significance.

### Summary of the OS and PS Analysis

The results indicated that for the *first key*, the effect of the PS was present in the two languages and for the two groups (i.e., participants processed high PS cognates faster than low PS cognates), although PS effects were stronger for the HSs than LBs. In LBs, the PS effect was stronger in Spanish than in English. OS had an effect in the Spanish block (i.e., participants processed high OS cognates faster than low OS cognates) but this effect interacted with PS. That is, the difference between

P+ and P– conditions was significant only for the high OS condition. In English, there was no effect of OS.

For the *rest of word*, the effect of PS in reaction time was restricted to Spanish in the HSs. However, in Spanish, it was present for accuracy (cognates with high PS had a better performance than cognates with low PS), but in English depended on OS (in cognates with high OS, there were no differences between P+ and P–. However, in low OS cognates, the accuracy was higher for P+ than P– cognates). Regarding the accuracy, the effect of OS depended on group and language, so that for the HSs, the effect appeared in Spanish, but not in English, whereas in the LBs, the effect was not evident. OS tended to be more influential for typing the rest of the word than the first key (which was more influenced by PS).

## DISCUSSION

The main goal of this study was to investigate language co-activation and the role of cognate status during bilingual writing using a writing to dictation task. More specifically, we investigated the relative contributions of the profile of participants' language backgrounds by testing two bilingual populations: LBs (L1: English; L2: Spanish) and HSs (Majority language: English; Minority language: Spanish) which were both immersed in an English dominant context but differed in the level of formal literacy received in Spanish. The main goal was to analyze performance during typing of cognate and non-cognate words and examine how different degrees of orthographic similarity (OS) and phonological similarity (PS) in cognates affected writing times and accuracy. Importantly, from a theoretical standpoint, it is not completely clear how non-identical but similar cognates are lexically represented, what the role of orthographic and phonological similarity is in shaping these representations, especially when bilingualism is modulated by more or less exposure to formal education in one language. Moreover, previous experiments on cognate similarity have used reading tasks with visual presentations which may have obscured the role of phonological similarity. Critically, here, we use a writing to dictation task in which words were orally presented but orthographically implemented, therefore providing a tool to unveil the role of both phonological and orthographic similarities. In addition, and very key to this study, the use of writing could also unveil possible differences in the nature of language co-activation for bilinguals with different language experiences. In the following subsections, we will discuss the reported results to examine the influence of cognate status, the impact of OS and PS in language co-activation, and the diversity of language and learning backgrounds on the current task.

### The Consequences of Co-activation in Writing to Dictation: The Overall Effect of Cognate Status

The results of our experiment shed some light on the nature of cognate effects during a writing to dictation task. Previous

studies have shown that cognates are “special” because they share more semantic, orthographic, and phonological characteristics between languages than non-cognates (Voga and Grainger, 2007). Cognate facilitation effects have been widely reported in bilinguals and reflect language co-activation in reading, visual word recognition (e.g., Lemhöfer and Dijkstra, 2004; Dijkstra et al., 2010; Peeters et al., 2013), and in translation (Muscalu and Smiley, 2018). In the present experiment, cognate effects were also modulated by the language experience of the bilingual and the language in which the writing task was performed. More specifically, cognate facilitation was only present in HSs while processing in the minority language (Spanish), providing evidence of co-activation with the majority language (English). However, the results demonstrated an unexpected cognate interference effect in English (L1/majority language) with cognates being less accurate and slower than non-cognates in both groups.

Although cognate interference is not a common finding, some previous studies have found a similar effect (Schwartz et al., 2007; Dijkstra et al., 2010; Comesaña et al., 2012; Muscalu and Smiley, 2018). Critically, in all of them, non-identical cognates were included as experimental material suggesting that the degree of OS and PS in cognates may have an important impact during word processing. The BAST (Tainturier, 2019) has proposed that the strength of co-activation is mediated by the degree of OS and PS between the two languages, so the relative proportion of high and low similarity cognates can modulate the resulting facilitation vs. interference effects. Importantly, in the present study, cognates with high orthographic and phonological similarity (O+P+) were intermixed with cognates with low OS or PS (O−P+ and O+P−) and cognates with low OS and PS (O−P−). The fact that low similarity cognates represented one-third of the cognate stimuli might have masked the expected cognate facilitation effect. Thus, cognates are generally expected to produce co-activation of the two languages, and in turn facilitation, but the salient change in the code/representation (orthographic or phonological) of non-identical cognates may have produced competition and impaired their processing. At this point, competition between the two language representations would trigger lateral inhibition in order to reduce interference and select the appropriate representation (for a similar interpretation see Comesaña et al., 2012). Because non-cognates produce much weaker between language co-activation than cognate words, competition between representations would also be weaker for non-cognates relative to cognates (even for low similarity cognates). The role of inhibition when selecting among lexical competitors has also been proposed by others (Borragan et al., 2018; Broersma et al., 2016; Filippi et al., 2014). In line with this interpretation, previous research has found larger error monitoring effects and higher recruitment of brain regions dedicated to control while processing non-identical cognates relative to control words (Declerck et al., 2017; Peeters et al., 2019).

In addition, our results showed that the interference effect was found in the L1/majority language in both LBs and HSs, replicating previous production studies which showed a reversed dominance effect, exemplified by more intrusion errors in the dominant language (Gollan et al., 2014; Gollan and Goldrick,

2016; Li and Gollan, 2018). In this direction, some studies have pointed out that language processing in the L1/majority language could be largely mediated by an automatic process of orthography to phonology conversions, while processing in the L2 is more attentionally demanding (Plat et al., 2018). We propose that the manipulated similarities and differences in phonology and orthography in the current study might have directly affected the phonology to orthography conversion (POC). Since the L1/majority language is mediated by automatic processes, it is easier to observe interference effects. On the contrary, during L2/minority language, processing is more demanding, and therefore, the interference effect is reduced. The fact that interference occurs for HSs in the majority language (English) even though Spanish is their L1 may suggest that the regulatory processes are dependent on language experience and proficiency.

## The Nature of Language Co-activation: The Role of PS and OS

The strength of language co-activation is mediated by the degree of orthographic and phonological similarity between languages (Tainturier, 2019). Nevertheless, orthographic processing has been the focus of most studies (e.g., Van Kesteren et al., 2012; Peeters et al., 2013; Casaponsa et al., 2014; Hoversten et al., 2017; Muscalu and Smiley, 2018), reporting in a general larger cognate facilitation effects with greater OS (Dijkstra et al., 2010). Crucially, cognates can also vary in the degree of phonological similarity (PS) across languages. However, the role of PS and the interaction of PS with OS have received little to no attention. Very few studies have explored the interplay of PS and OS during word processing and most have relied on a reading task in which orthographic processing is imperative (Schwartz et al., 2007; Comesaña et al., 2012). For example, previous studies have demonstrated that the positive effect of PS (i.e., faster RTs for cognates with high PS than cognates with low PS) was mediated by the OS (Schwartz et al., 2007; Comesaña et al., 2012). In those studies, the PS effects only emerged in high OS conditions (i.e., the response in O+P+ condition was faster than the responses in O+P− condition). However, there were no differences between high and low PS in cognates with low OS (there were no differences between O−P+ and O−P−). In other words, if common orthographic L1/L2 nodes map into different phonological L1/L2 nodes, it can create confusion, slowing down the processing of the word (Doctor and Klein, 1992; Dijkstra et al., 1999; Dijkstra and Van Heuven, 2002; Schwartz et al., 2007).

The absence of PS effects in low OS conditions reported in previous studies has been explained by the orthographic autonomy hypothesis which proposes that written production is not dependent on spoken production and therefore not dependent on phonological information (Rapp and Caramazza, 1997). In reading, orthographic retrieval is mandatory, and the co-activation of language nodes would be mediated by OS. In addition, in the O− condition, the co-activated representations compete for selection, and inhibition would

be triggered to achieve successful processing (in the Comesaña and colleagues' and in the Schwartz and colleagues' studies). This first orthographic filter would reduce the spread of activation to phonology (see **Figure 1** in which the arrows on the left represent the reduction of spread in the O-condition). As mentioned, phonological processing in reading is delayed with respect to orthographic processing because the stimuli are visually presented and mapping of orthography to phonology only occurs after orthographic analyses have taken place. However, writing production paradigms, and especially writing to dictation tasks, can be key to study the role of phonology because these tasks involve phonological input and orthographically oriented responses, such that phonological processing is mandatory (obligatory phonological mediation hypothesis; Geschwind, 1969).

Contrary to previous studies, the results of our writing to dictation task showed a general PS effect in the first key latency and accuracy in most conditions of the experiment. In the first key latency analysis, the PS effect (faster RTs for cognates with high PS than cognates with low PS) was present for both groups (LBs and HSs) in English (L1/majority language) and Spanish (L2/minority language), suggesting primacy of phonological processing facilitating the access to the lexical representations of the words. The first filter would therefore be phonological, so in the low OS condition, phonological information would continue to be processed, because the first filter, in this case, did not reduce the spread of activation to phonology (see **Figure 1**, specifically see the arrows on the far right).

In contrast, PS effects for the rest of the word, although present in accuracy, were not present in LBs and it interacted with OS in HSs (in Spanish). This pattern suggests that the role of phonology is smaller as the time course progresses and the influence of orthography gains relevance. The fact that the OS effect was more consistently found in Spanish than in English in the rest of the word analyses suggests that the way the words are processed in each language could be different (i.e., after the first key). Performance on the rest of the word in the writing task has been attributed to sublexical processing (Muscalu and Smiley, 2018; Iniesta et al., 2021). Dual-route theories of reading propose that transparent orthographies, such as Spanish, rely on phoneme to grapheme processing, contrary to deeper languages, such as English, which uses direct access to lexical representations (orthographic depth hypothesis; Frost, 1994, 2012). So, the OS, which is a sublexical characteristic, would more directly affect sublexical processing (the POC system) than lexical processing explaining the greater role of OS in Spanish.

In sum, differences in the time course of orthographic and phonological activation during reading vs. writing to dictation tasks explain the differences in the impact of OS and PS. The bilingual interactive activation BIA+ model (Dijkstra and Van Heuven, 2002) introduces the "temporal delay assumption" to explain that under some conditions, cross-linguistic phonological, orthographic, and/or semantic effects may be absent due to task demands. Reading requires orthographic activation prior to phonological activation, and therefore, the late phonological

activation would not affect response times (Brysbaert et al., 2002). However, during a writing to dictation task, the phonological processing precedes activation of orthographic information, and therefore, the phonology may directly impact the performance. The fact that phonological processing occurs early in writing to dictation explain the generalized PS effects in all experimental conditions (faster responses for cognates with high PS than for cognates with low PS).

Following previous studies, we decided to use OS and PS as dichotomous variables to directly compare reading and writing to dictation (Schwartz et al., 2007; Comesaña et al., 2012). However, the threshold used to classify cognates as high or low similarity is somewhat arbitrary, and future research in this field should consider OS and PS as continuous variables.

## The Role of the Learning Environment in Language Co-activation

In our experiment, we included two groups of bilinguals: LBs and HSs. We hypothesized that the relationship between the L1 and the L2 could be influenced by their linguistic learning background (Kroll et al., 2006; Dijkstra et al., 2010). More specifically, differences in literacy and exposure to writing and reading between the two groups might modulate the co-activation effects and the relative roles of OS and PS in L1 and L2 processing. The two groups did not differ in the LEAP-Q measures for English (L1/majority language): There were no differences in the years of schooling in an English context, nor in their self-assessed language skills for speaking, understanding, and reading, nor in their reading contribution to learning measure which reflects L1 formal learning and regulates learning at school (Iniesta et al., 2021). For Spanish, however (L2/minority language), the LEAP-Q highlighted significant differences in the self-assessed skills in speaking and understanding with HSs scoring higher than LBs. Critically, in skills that were more related to formal language use, like reading, there were no group differences, and additionally, HSs showed a lower score for the score reading contribution to learning. In addition, and in accordance with previous studies (Elola and Mikulski, 2016), scores in the Spanish tests showed worse performance for HSs than LBs (PROESC and TECLE). Hence, even though the years of exposure to Spanish were greater in the HSs, they showed more orthographic difficulties in Spanish than the LBs, presumably due to the fact that their input during learning was mainly phonological, resulting in a less accumulated literacy practice (see the weaker links hypothesis; Gollan et al., 2008).

In the same direction of PROESC and TECLE, the HSs showed worse performance in the writing to dictation task (relative to LBs), specifically in the latency of the rest of word performance. This suggests that the HSs might have greater difficulties in sublexical processing, where the orthographic form retrieval is especially important. In addition, analysis of the RTs showed an interaction between Group and Language. This interaction indicated that both groups were faster in

English than Spanish, but the magnitude of the difference was greater for the LBs than for the HSs when looking at the first key performance (lexical access), and the magnitude of the difference was greater for the HSs than the LBs when considering the rest of the word performance (sublexical processing). Again, this pattern suggests that the HSs might have more difficulties retrieving the word form in both languages (English and Spanish) although these difficulties become more evident during writing production in the minority language, presumably due to less accumulated practice as a result of their learning background (see also Gollan et al., 2008).

Regarding the OS and PS between languages, there were subtle differences between groups. In the RTs analysis of the first key (lexical) latency, the results showed a Group  $\times$  Language  $\times$  PS interaction. Even though there were significant differences between high PS and low PS in both groups and both languages (Spanish and English), the magnitude of the difference was higher in Spanish in the LBs (cognates with high PS were typed faster than cognates with low PS). A possible interpretation of this effect is that when LBs type the first key in Spanish, the English phonology is more co-activated than when they are typing in English and Spanish is co-activated. In contrast, for HSs, there were no magnitude differences between languages for phonology. There were no accuracy differences while processing O+ and O- cognates for the first key, suggesting that for HSs the sensitivity to the OS is reduced in both languages. This pattern supports previous studies that show phonological advantages for HSs relative to LBs (Chang et al., 2011; Gor, 2014), but also orthographic disadvantages (Elola and Mikulski, 2016).

In sum, these results add to the current literature on bilingual language co-activation by demonstrating that the language learning environment, especially formal exposure to reading and writing in a given language, can not only modulate proficiency but also affect how the languages are co-activated and how they interact.

## CONCLUSION

The present study provides evidence that language co-activation during writing production in L1 and L2 is modulated by OS, but also, and more important, by PS across languages. Writing to dictation involves phonology from the very early processing stages so that PS contributes to facilitating access to the lexical representation of the words. Hence, contrary to previous studies on reading, the PS effects were very pervasive during lexical access (first key latency, [i.e., participants process cognates with high PS faster than cognates with low PS]), although they showed modulation with orthography during the implementation of writing while typing the rest of the word (sublexical processing). In contrast, the effect of OS was not extensively evident during lexical access (first key), and it had a more important role during the sublexical processing (rest of the word). In addition, the results provide evidence about the impact of literacy differences for orthographic and phonological co-activation during writing

production (in this case, the acquisition of the Spanish L2/minority language).

To conclude, the interplay of OS and PS underlying cross-linguistic influence in bilinguals seems to be dependent on the relative order in which orthographic and phonological processing occur, and this pattern can be modulated by the task that bilinguals are performing and by the language learning environment of the bilinguals. Commonly, bilingual competence is conceptualized as a continuum. In this continuum, the study of HSs is especially important because it allows for an exploration of how different cultural, linguistic, and educational contexts influence language learning and the relationship between languages.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Florida Institutional Review Board (IRB): protocol #2019-02427. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AI, DP, and TB contributed the conception and design of the study. AI and ER organized the data collection. AI and DP performed the statistical analysis. AI wrote the first draft of the manuscript. All authors wrote sections of the manuscript and contributed to the manuscript revision, read, and approved the submitted version.

## FUNDING

The current research was completed and thanks to financial aid provided by the doctoral research grant FPU16/01748 to AI and grants from the Ministerio de Ciencia, Innovación y Universidades-Fondos Feder to TB (PGC2018-093786-B-I00) and DP (PCIN-2015-165-C02-01 and PSI2017-89324-C2-1-P) and from the Feder Andalucía to TB (A-CTS-111-UGR18 and P20.00107).

## ACKNOWLEDGMENTS

We wish to acknowledge Jorge Valdés Kroff and Mayori Matos for their help in stimuli recording and the Department of Spanish and Portuguese Studies at the University of Florida for their help in recruiting participants.

# REFERENCES

- Alonso, M., Díez, E., and Fernandez, A. (2015). Subjective age-of-acquisition norms for 4.640 verbs in Spanish. *Behav. Res. Methods* 48, 1337–1342. doi: 10.3758/s13428-015-0675-z
- Anderson, J. A. E., Mak, L., Keyvani Chahi, A., and Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behav. Res. Methods* 50, 250–263. doi: 10.3758/s13428-017-0867-9
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using LME4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theo. Ling.* 39, 129–181. doi: 10.1515/tl-2013-0009
- Boersma, P., and Weenink, D. (2012). *Praat: Doing phonetics by computer [Computer software]*. Amsterdam: Institute of Phonetic Sciences.
- Borragan, M., Martin, C. D., De Bruin, A., and Duñabeitia, J. A. (2018). Exploring different types of inhibition during bilingual language production. *Front. Psychol.* 9:2256. doi: 10.3389/fpsyg.2018.02256
- Bowers, J. S., Mimouni, Z., and Arguin, M. (2000). Orthography plays a critical role in cognate priming: evidence from french/english and arabic/french cognates. *Mem. Cogn.* 28, 1289–1296. doi: 10.3758/BF03211829
- Broersma, M., Carter, D., and Acheson, D. J. (2016). Cognate costs in bilingual speech production: Evidence from language switching. *Front. Psychol.* 7:1461. doi: 10.3389/fpsyg.2016.01461
- Brysbaert, M., and Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *J. Cogn.* 1:9. doi: 10.5334/joc.10
- Brysbaert, M., Van Wijnendaele, I., and Duyck, W. (2002). On the temporal delay assumption and the impact of non-linguistic context effects. *Biling. Lang. Cogn.* 5, 199–201. doi: 10.1017/S1366728902213012
- Brysbaert, M., Warriner, A., and Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behav. Res. Methods* 46, 904–911. doi: 10.3758/s13428-013-0403-5
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Biling. Lang. Cogn.* 16, 32–48. doi: 10.1017/S1366728912000120
- Carrasco-Ortiz, H., Amengual, M., and Gries, S. T. (2019). Cross-language effects of phonological and orthographic similarity in cognate word recognition: the role of language dominance. *Ling. Approaches Bilingualism* 11, 1–29. doi: 10.1075/lab.18095.car
- Casaponsa, A., Carreiras, M., and Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: the impact of orthographic markedness. *Front. Psychol.* 5:424. doi: 10.3389/fpsyg.2014.00424
- Chacón, T. (1997). *Ortografía Española*. Madrid: UNED.
- Chang, C. B., Yao, Y., Haynes, E. F., and Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *J. Acoust. Soc. Am.* 129, 3964–3980. doi: 10.1121/1.3569736
- Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Res.* 1147, 192–208. doi: 10.1016/j.brainres.2007.01.137
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295X.108.1.204
- Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., and García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 614–635. doi: 10.1037/xlm0000065
- Comesaña, M., Sánchez-Casas, R., Soares, A. P., Pinheiro, A. P., Rauber, A., Frade, S., et al. (2012). The interplay of phonology and orthography in visual cognate word recognition: An ERP study. *Neurosci. Lett.* 529, 75–79. doi: 10.1016/j.neulet.2012.09.010
- Costa, A., Caramazza, A., and Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1283–1296. doi: 10.1037//0278-7393.26.5.1283
- Costa, A., Santesteban, M., and Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain Lang.* 94, 94–103. doi: 10.1016/j.bandl.2004.12.002
- Costa, A., and Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709
- Cuetos, F., Ramos, J., and Ruano, E. (2002). *PROESC. Bateria De Evaluación De Los Procesos De Escritura*. Madrid: TEA.
- Declerck, M., Lemhöfer, K., and Grainger, J. (2017). Bilingual language interference initiates error detection: Evidence from language intrusions. *Biling. Lang. Cogn.* 20, 1010–1016. doi: 10.1017/S1366728916000845
- Dijkstra, T., Grainger, J., and van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *J. Mem. Lang.* 41, 496–518. doi: 10.1006/jmla.1999.2654
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *J. Mem. Lang.* 62, 284–301. doi: 10.1016/j.jml.2009.12.003
- Dijkstra, T., and Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Biling. Lang. Cogn.* 5, 175–197. doi: 10.1017/S1366728902003012
- Doctor, E. A., and Klein, D. (1992). “Phonological processing in bilingual word recognition” in *Cognitive Processing in Bilinguals*. ed. R. J. Harris (Amsterdam, Netherlands: Elsevier Science), 237–252.
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., and Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behav. Res. Methods* 45, 1246–1258. doi: 10.3758/s13428-013-0326-1
- Elola, I., and Mikulski, A. M. (2016). Similar and/or different writing processes? A study of Spanish foreign language and heritage language learners. *Hispania* 99, 87–102. doi: 10.1353/hpn.2016.0000
- Filippi, R., Karaminis, T., and Thomas, M. S. C. (2014). Language switching in bilingual production: empirical data and computational modelling. *Biling. Lang. Cogn.* 17, 294–315. doi: 10.1017/S1366728913000485
- Fricke, M., Zirnstein, M., Navarro-Torres, C. A., and Kroll, J. F. (2019). Bilingualism reveals fundamental variation in language processing. *Biling. Lang. Cogn.* 22, 200–207. doi: 10.1017/S1366728918000482
- Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *J. Exp. Psychol. Learn. Mem. Cogn.* 20:116. doi: 10.1037//0278-7393.20.1.116
- Frost, R. (2012). Towards a universal model of reading. *Behav. Brain Sci.* 35, 263–279. doi: 10.1017/S0140525X11001841
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Biling. Lang. Cogn.* 12, 213–237. doi: 10.1017/S1366728909004015
- Geschwind, N. (1969). “Problems in the anatomical understanding of the aphasia,” in *Contributions to Clinical Neuropsychology*. ed. A. L. Benton (Chicago: Aldine), 107–128.
- Gollan, T. H., and Goldrick, M. (2016). Grammatical constraints on language switching: Language control is not just executive control. *J. Mem. Lang.* 90, 177–199. doi: 10.1016/j.jml.2016.04.002
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., and Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory Cogn.* 33, 1220–1234. doi: 10.3758/BF03193224
- Gollan, T. H., Schotter, E. R., Gomez, J., Murillo, M., and Rayner, K. (2014). Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychol. Sci.* 25, 585–595. doi: 10.1177/0956797613512661
- Gor, K. (2014). Raspberry, not a car: context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Front. Psychol.* 5:1449. doi: 10.3389/fpsyg.2014.01449
- Green, D. W., and Abutalebi, J. (2015). Language control and the neuroanatomy of bilingualism: In praise of variety. *Lang. Cogn. Neurosci.* 31, 340–344. doi: 10.1080/23273798.2015.1084428

- Guasch, M., Boada, R., Ferré, P., and Sánchez-Casas, R. (2013). NIM: A Web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behav. Res. Methods* 45, 765–771. doi: 10.3758/s13428-012-0296-8
- Guasch, M., Ferré, P., and Haro, J. (2017). Pupil dilation is sensitive to the cognate status of words: further evidence for non-selectivity in bilingual lexical access. *Biling. Lang. Cogn.* 20, 49–54. doi: 10.1017/S1366728916000651
- Halekoh, U., and Højsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbrtest. *J. Stat. Softw.* 59, 1–30. doi: 10.18637/jss.v059.i09
- Hoshino, N., and Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition* 106, 501–511. doi: 10.1016/j.cognition.2007.02.001
- Houghton, G., and Forzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cogn. Neuropsychol.* 20, 115–162. doi: 10.1080/02643290242000871
- Hoversten, L. J., Brothers, T., Swaab, T. Y., and Traxler, M. J. (2017). Early processing of orthographic language membership information in bilingual visual word recognition: Evidence from ERPs. *Neuropsychologia* 103, 183–190. doi: 10.1016/j.neuropsychologia.2017.07.026
- Hyltenstam, K., and Abrahamsson, N. (2003). “Age of onset and ultimate attainment in near-native speakers of Swedish as a second language,” in *Multilingualism in Global and Local Perspectives*. eds. K. Fraurud and K. Hyltenstam (Sweden: Centre for Research on Bilingualism and Rinkeby Institute of Multilingual Research, Stockholm University), 319–340.
- Iniesta, A., Paolieri, D., Serrano, F., and Bajo, M. T. (2021). *Bilingual Writing Co-Activation: Lexical and Sublexical Processing in A Word Dictation Task*. Bilingualism: Language and Cognition, 1–16.
- Ivanova, I., and Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta. Psychol.* 127, 277–288. doi: 10.1016/j.actpsy.2007.06.003
- Kroll, J. F., Bobb, S. C., and Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Biling. Lang. Cogn.* 9, 119–135. doi: 10.1017/S1366728906002483
- Kroll, J. F., Dussias, P. E., and Bajo, M. T. (2018). Language use across international contexts: Shaping the minds of L2 speakers. *Annu. Rev. Appl. Linguist.* 38, 60–79. doi: 10.1017/S0267190518000119
- Kroll, J. F., Gullifer, J. W., and Rossi, E. (2013). The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two or more languages. *Annu. Rev. Appl. Linguist.* 33:102. doi: 10.1017/S0267190513000111
- Kroll, J., van Hell, J., Tokowicz, N., and Green, D. (2010). The revised hierarchical model: A critical review and assessment. *Biling. Lang. Cogn.* 13, 373–381. doi: 10.1017/S136672891000009X
- Kuperman, V., Stadthagen-Gonzalez, H., and Brysbaert, M. (2012). Erratum to: Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* 45:900. doi: 10.3758/s13428-013-0348-8
- Lemhöfer, K., and Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Mem. Cogn.* 32, 533–550. doi: 10.3758/BF03195845
- Lemhöfer, K., Spalek, K., and Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *J. Mem. Lang.* 59, 312–330. doi: 10.1016/j.jml.2008.06.005
- Li, C., and Gollan, T. H. (2018). Cognates interfere with language selection but enhance monitoring in connected speech. *Mem. Cogn.* 46, 923–939. doi: 10.3758/s13421-018-0812-x
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502. doi: 10.3758/s13428-016-0809-y
- Macizo, P. (2016). Phonological coactivation in the bilinguals’ two languages: Evidence from the color naming task. *Biling. Lang. Cogn.* 19, 361–375. doi: 10.1017/S136672891500005X
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230. doi: 10.1371/journal.pone.0043230
- Marian, V., Blumenfeld, H., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)
- Marian, V., and Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Appl. Psycholinguist.* 24, 173–193. doi: 10.1017/S0142716403000092
- Marín, J., and Carrillo, M. (1999). *Test Colectivo De Eficacia Lectora (TECLE)*. Spain: Universidad de Murcia.
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism. Re-examining the Age Factor*. Amsterdam: John Benjamins.
- Montrul, S. (2013). *El bilingüismo En El Mundo Hispanohablante*. United States: John Wiley and Sons.
- Muscalu, L., and Smiley, P. (2018). The illusory benefit of cognates: Lexical facilitation followed by sublexical interference in a word typing task. *Biling. Lang. Cogn.* 22, 848–865. doi: 10.1017/s1366728918000792
- Paradis, J. (2004). The relevance of specific language impairment in understanding the role of transfer in second language acquisition. *Appl. Psycholinguist.* 25, 67–82. doi: 10.1017/S0142716404001043
- Peeters, D., Dijkstra, T., and Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *J. Mem. Lang.* 68, 315–332. doi: 10.1016/j.jml.2012.12.003
- Peeters, D., Vanlangendonck, F., Rueschemeyer, S. A., and Dijkstra, T. (2019). Activation of the language control network in bilingual visual word recognition. *Cortex* 111, 63–73. doi: 10.1016/j.cortex.2018.10.012
- Place, S., and Hoff, E. (2011). Properties of dual language exposure that influence two-year-olds’ bilingual proficiency. *Child Dev.* 82, 1834–1849. doi: 10.1111/j.1467-8624.2011.01660.x
- Plat, R., Lowie, W., and de Bot, K. (2018). Word naming in the L1 and L2: A dynamic perspective on automatization and the degree of semantic involvement in naming. *Front. Psychol.* 8:2256. doi: 10.3389/fpsyg.2017.02256
- Polinsky, M. (2011). Reanalysis in adult heritage language: A case for attrition. *Stud. Second. Lang. Acquis.* 33, 305–328. doi: 10.1017/S027226311000077X
- Purcell, J., Turkeltaub, P. E., Eden, G. F., and Rapp, B. (2011). Examining the central and peripheral processes of written word production through meta-analysis. *Front. Psychol.* 2:239. doi: 10.3389/fpsyg.2011.00239
- Rapp, B., and Caramazza, A. (1997). From graphemes to abstract letter shapes: Levels of representation in written spelling. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1130–1152. doi: 10.1037//0096-1523.23.4.1130
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rothman, J., and Guijarro-Fuentes, P. (2010). Input quality matters: Some comments on input type and age-effects in adult SLA. *Appl. Linguist.* 31, 301–306. doi: 10.1093/applin/amq004
- Schwartz, A. I., Kroll, J. F., and Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Lang. Cogn. Process.* 22, 106–129. doi: 10.1080/01690960500463920
- Tainturier, M.-J. (2019). “A theory of bilingual spelling in alphabetic systems,” in *Spelling and Writing Words: Theoretical and Methodological Advances*. Vol. 39. eds. T. Olive and C. Perret (Netherlands: Brill)
- Valdés Kroff, J. R., Rooijakkers, F., and Parafita, M. C. (2019). Spanish grammatical gender interference in papiamentu. *Languages* 4:78. doi: 10.3390/languages4040078
- van Hell, J. G., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychon. Bull. Rev.* 9, 780–789. doi: 10.3758/BF03196335
- van Heuven, W. J. B., Dijkstra, T., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Mem. Lang.* 39, 458–483. doi: 10.1006/jmla.1998.2584
- Van Kesteren, R., Dijkstra, T., and De Smedt, K. (2012). Markedness effects in norwegian-english bilinguals: task-dependent use of language-specific letters and bigrams. *Q. J. Exp. Psychol.* 65, 2129–2154. doi: 10.1080/17470218.2012.679946
- Van Orden, G. C., Johnston, J. C., and Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 371–386. doi: 10.1037//0278-7393.14.3.371
- Voga, M., and Grainger, J. (2007). Cognate status and cross-script translation priming. *Mem. Cogn.* 35, 938–952. doi: 10.3758/BF03193467
- Von Studnitz, R. E., and Green, D. W. (2002). The cost of switching language in a semantic categorization task. *Biling. Lang. Cogn.* 5, 241–251. doi: 10.1017/S1366728902003036
- Woutersen, M., de Bot, K., and Weltens, B. (1995). The bilingual lexicon: Modality effects in processing. *J. Psycholinguist. Res.* 24, 289–298. doi: 10.1007/BF02145058

Yates, M., Locker, L., and Simpson, G. B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Mem. Cogn.* 31, 856–866. doi: 10.3758/BF03196440

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Iniesta, Rossi, Bajo and Paolieri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Does L2 Proficiency Impact L2-L1 Transfer While Reading L1 Collocations? Evidence From Behavioral and ERP Data

Agnieszka Otwinowska<sup>1\*</sup>, Marta Marecka<sup>2</sup>, Alba Casado<sup>2</sup>, Joanna Durlík<sup>2</sup>, Jakub Szewczyk<sup>2,3</sup>, Marcin Opacki<sup>1</sup> and Zofia Wodniecka<sup>2\*</sup>

## OPEN ACCESS

### Edited by:

Cristina Cacciari,  
University of Modena and Reggio  
Emilia, Italy

### Reviewed by:

Sybrine Bultena,  
Radboud University Nijmegen,  
Netherlands  
Anna Hatzidaki,  
National and Kapodistrian University  
of Athens, Greece

### \*Correspondence:

Agnieszka Otwinowska  
a.otwinowska@uw.edu.pl  
Zofia Wodniecka  
zofia.wodniecka@uj.edu.pl

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 February 2021

**Accepted:** 23 August 2021

**Published:** 29 September 2021

### Citation:

Otwinowska A, Marecka M,  
Casado A, Durlík J, Szewczyk J,  
Opacki M and Wodniecka Z (2021)  
Does L2 Proficiency Impact L2-L1  
Transfer While Reading L1  
Collocations? Evidence From  
Behavioral and ERP Data.  
Front. Psychol. 12:673761.  
doi: 10.3389/fpsyg.2021.673761

<sup>1</sup> Faculty of Modern Languages, Institute of English Studies, University of Warsaw, Warsaw, Poland, <sup>2</sup> Faculty of Philosophy, Institute of Psychology, Jagiellonian University, Kraków, Poland, <sup>3</sup> Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Multi-word expressions (MWEs) are fixed, conventional phrases often used by native speakers of a given language (L1). The type of MWEs investigated in this study were collocations. For bilinguals who have intensive contact with the second language (L2), collocational patterns can be transferred from the L2 to the L1 as a result of cross-linguistic influence (CLI). For example, bilingual migrants can accept collocations from their L2 translated to their L1 as correct. In this study, we asked whether such CLI is possible in native speakers living in the L1 environment and whether it depends on their L2 English proficiency. To this end, we created three lists of expressions in Polish: (1) well-formed Polish verb-noun collocations (e.g., *ma sens* – \*has sense), (2) collocational calques from English (loan translations), where the English verb was replaced by a Polish translation equivalent (e.g., *\*robi sens* – makes sense), and, as a reference (3) absurd verb-noun expression, where the verb did not collocate with the noun (e.g., *\*zjada sens* – \*eats sense). We embedded the three types of collocations in sentences and presented them to L1 Polish participants of varying L2 English proficiency in two experiments. We investigated whether L2 calques would (1) be explicitly judged as non-native in the L1; (2) whether they would evoke differential brain response than native L1 Polish equivalents in the event-related potentials (ERPs). We also explored whether the sensitivity to CLI in calques depended on participants' level of proficiency in L2 English. The results indicated that native speakers of Polish assessed the calques from English as less acceptable than the correct Polish collocations. Still, there was no difference in online processing of correct and calques collocations as measured by the ERPs. This suggests a dissociation between explicit offline judgments and indices of online language processing. Interestingly, English L2 proficiency did not modulate these

effects. The results indicate that the influence of English on Polish is so pervasive that collocational calques from this language are likely to become accepted and used by Poles.

**Keywords:** L2-L1 transfer, multi-word expression, collocation, ERP, acceptability judgments, cross-linguistic influence

## INTRODUCTION

In all languages, certain words co-occur and form fixed sequences called multiword expressions (MWEs, Siyanova-Chanturia, 2013, 2015), such as collocations (e.g., *take a picture*). For learners of second or foreign languages (L2) word choices in L2 MWEs may sometimes be arbitrary (Szudarski and Conklin, 2014), but native speakers of a language (L1) can judge that “something is wrong” when certain words in the MWE are replaced by other words. For example, when native speakers of English are asked to make an explicit judgment on collocations, most would prefer the phrase *take a picture* to the phrase *make a picture*. In a similar context, a native speaker of Polish would prefer *zrobić zdjęcie* (make a picture) to *wziąć zdjęcie* (take a picture, which in Polish implies taking the picture in one's hands). However, while native speakers might have explicit, metalinguistic opinions about MWE use, are they equally sensitive to such subtle differences as *make/take a picture* when processing collocations?

Evidence suggests that for bilinguals living in an L2 environment, explicit judgments between native and non-native collocations might become blurred (Laufer, 2003; Schmid and Köpke, 2017). This is due to the influence of the L2 on the L1 (Jarvis and Pavlenko, 2008), known as cross-linguistic influence (CLI). In our study, we asked whether the native and non-native distinctions between collocations can become blurred due to CLI in the case of those native speakers who live in the L1 environment and who use English as the “global language” (Seidlhofer, 2013). We presented L1 Polish participants of varying L2 English proficiency with sentences containing Polish collocations, as well as sentences where those collocations were replaced by their calques (loan translations) from English. We investigated whether such calques would (1) be explicitly judged as non-native and (2) evoke differential brain response than their native Polish equivalents in the event-related potentials (ERPs). We also explored whether sensitivity to CLI in calques depends on participants' level of proficiency in L2 English.

## What Are Multi-Word Expressions and How to Identify Them?

To express a specific meaning, native speakers of any language often use multiword sequences, fixed expressions or phrases characterized by a degree of connectedness and recognized as conventionalized (Siyanova-Chanturia, 2013). Those multiword sequences can be called formulaic expressions and formulaic sequences (Wray and Perkins, 2000; Kecskes, 2015; Carroll and Conklin, 2020), multiword items (Siyanova-Chanturia and Omidian, 2019), or multi-word expressions (MWEs, Siyanova-Chanturia, 2013). MWEs are defined as combinations of words for which syntactic or semantic properties of the expression cannot be obtained from their component parts (Sag et al.,

2002). There are many types of MWEs, including grammatical expressions (*is going to*), phrasal verbs (*look up*), situation bound utterances (*How can I help you?*), binomial expressions (*bread and butter*), idioms (*take the bull by the horns*), and collocations (*take a picture*; Wray, 2005; Kecskes, 2015; Siyanova-Chanturia, 2015). Although MWEs differ in terms of their fixedness (how restricted the word combination is) and compositionality (the degree to which the meaning of the MWE stems from the meaning of the separate words), they can be identified based on their recurrence in the given language. This means that in natural language MWEs recur more frequently than comparable phrases that are less fixed (Carroll and Conklin, 2020).

Based on the characteristics of fixedness and recurrence, two predominant methods of identifying MWEs have been proposed. The “phraseological approach” uses native speaker intuitions in the assessment of how fixed or non-compositional particular word combinations are with respect to their meaning (Siyanova-Chanturia, 2015; Siyanova-Chanturia and Omidian, 2019). A more quantitative method based on language corpora, is the “frequency-based approach” (Sinclair, 1991; Durrant and Schmitt, 2009). It investigates the relationship of words that co-occur within texts with greater than random probability. In a language corpus, two or more words are considered to be a MWE if they co-occur more often than predicted from the frequencies of the separate words. For instance, collocations (such as, *take a picture*) are pairs of words that co-occur more frequently than we would expect by chance (Sinclair, 1991; Carroll and Conklin, 2020) and are compositional, so their meaning depends on the meanings of the particular words. However, the words in a collocation will be more strongly associated than words in other word combinations (*take a picture* vs. *\*make a picture*). Thus, statistical metrics of association strength are employed in the frequency-based approach to identifying MWEs. Most notably, they include the t-score, which is the confidence with which we can assert that there is an association between two words, and the mutual information (MI) score<sup>1</sup>, which is the strength of co-occurrence between two words that form an MWE (Sinclair, 1991; Siyanova-Chanturia and Omidian, 2019;

<sup>1</sup>The Mutual Information score is a measure of the strength of association between words *x* and *y*. It expresses the extent to which the observed frequency of co-occurrence differs from what is expected based on chance. In a given corpus, the MI is calculated on the basis of the number of times a pair of items was observed together, versus the number of times each of the items that form the pair were observed separately. MI is more likely to give high scores to fixed phrases whereas the t-score will yield significant collocates that occur relatively frequently. The t-score measures the confidence with which we can assert that there is an association (Harper, 2008). While the MI score is more sensitive to conventionally lexicalized collocations (e.g., *post graduate*) whose internal collocates are unlikely to occur as separate words within a particular corpus, the t-score, by taking corpus size into account, is more sensitive to situationally conditioned collocations (i.e., ones specific to a given corpus) such as “rely on” or “in the form of the,” which are not conventionalized, but may recur in a particular specialized corpus.

Carrol and Conklin, 2020). Because both approaches to identifying MWEs (frequency-based and phraseological, intuition based) have their own advantages and limitations (see Wray, 2005; Durrant and Schmitt, 2009), and are often employed in a complementary fashion (Siyanova-Chanturia and Omidian, 2019), both will also be used in the current study when creating experimental materials.

## MWEs in L1 Use and Processing

Multi-word expressions are commonly used by native speakers, especially in speech, and the frequency of MWEs in a native speaker's lexicon is almost equivalent to that of single words (Wray and Perkins, 2000). Linguistic literature suggests that native speakers use MWEs mainly for reasons of speech economy, because they are stored in memory as prefabricated "language chunks," are selected as wholes for production, and do not need to be decomposed to access their meaning in comprehension (Sinclair, 1991; Wray, 2005, 2012). Indeed, native speakers process MWEs such as idioms, binominals, and collocations faster than other word combinations. This manifests itself in faster reaction times in behavioral tasks and reading (e.g., Arnon and Snider, 2010; Vespignani et al., 2010; Rommers et al., 2013; Vilkaitė, 2016; Vilkaitė and Schmitt, 2019; Carrol and Conklin, 2020). For instance, Carrol and Conklin (2020) used eyetracking to compare reading times of three types of MWE, that is idioms, binominals, and collocations relative to control phrases. The results showed that native speakers of English read all three MWE types embedded in sentences faster than control phrases. Also Vilkaitė (2016) found that English native speakers read verb-noun collocations (e.g., *provide information*) embedded in sentences more quickly than control phrases (e.g., *compare information*), even in non-adjacent configurations separated by three other words (e.g., *provide some of the information*). Vilkaitė interpreted this as evidence against the suggestions that collocations and other MWEs are stored and processed as wholes and remain unanalyzed (Wray, 2005, 2012). Rather, what these data imply is that parts of a MWE are more predictable than other word combinations occurring in discourse.

The mechanism of prediction underlying discourse comprehension is based on the interplay between the language information already processed by our brain and the information currently being processed (Vespignani et al., 2010). Thanks to the previously processed linguistic information, the memory representation of a given word becomes activated even before this word occurs in the input directed to the person hearing or reading that discourse stretch (Szewczyk and Schriefers, 2018). If the sequence or co-occurrence of some words is highly predictable, the words previously processed activate the words to occur. Some studies into predictive effects in language comprehension have investigated such MWEs as idioms (e.g., Vespignani et al., 2010; Rommers et al., 2013), binominals and collocations (Siyanova-Chanturia et al., 2017; Carrol and Conklin, 2020). Their results point out that, indeed, parts of a MWE are highly predictable, most likely because they are composed of words that often co-occur in fixed patterns. Thus, the activation of one element of an MWE activates the other elements faster than in the case of less fixed

phrases, leading to faster processing of the entire expression (Arnon and Snider, 2010).

The predictability of MWEs when processed by native speakers can also be detected in the presence of specific components measured in the scalp-recorded event-related potentials (ERPs) used in the study of language comprehension. One of those components is the N400 (Kutas and Hillyard, 1980), which is a broad negative deflection that begins 200–300 ms after a word has been presented and reaches its peak after approximately 400 ms. The N400 always occurs when a native speaker of a language is perceiving any stimuli that are conceptually meaningful. It reflects the degree to which the conceptual representations associated with the stimulus have already been active in long-term memory. Accessing the representations that are already active leads to less negative N400 amplitudes than accessing stimuli that are less activated in memory. Comprehending a coherent sentence often leads to a pre-activation of meanings that are likely to occur in the upcoming part of a sentence. Thus, words that are congruent with the sentence often elicit N400 components with a less negative amplitude, but reading words incongruent with the prediction gives rise to a large N400 component. In other words, it is a typical finding that the amplitude of the N400 correlates with the extent to which a word's meaning is congruent with the meaning of the preceding context (Kutas and Federmeier, 2011; Szewczyk and Schriefers, 2018).

Although ERP studies on MWE processing are still limited, they all show that once the reader encounters an MWE and reads the first word, the predictability of the words in the remaining part of the MWE greatly increases. Cases when native speakers process MWEs in comparison with other, non-fixed expressions yield the less negative N400 amplitudes. This has been demonstrated for the processing of idioms in Dutch and Italian (e.g., *cry over spilt milk* vs. *cry over spilt coffee*; Vespignani et al., 2010; Rommers et al., 2013), binominal phrases in English (e.g., *knife and fork* vs. *spoon and fork*; Siyanova-Chanturia et al., 2017) and collocations in Spanish (*quite the opposite* vs. *all the opposite*; Molinaro and Carreiras, 2010). The reduced N400 when processing MWEs relative to the processing of other non-fixed word combinations indicates that the parts of an MWE were highly expected, possibly because their structure and meaning were stored in memory.

Yet another component reported in several ERP studies on MWEs is the P300, a positivity occurring in the 250–350 ms time-window on parietal electrodes (Molinaro and Carreiras, 2010; Vespignani et al., 2010; Siyanova-Chanturia et al., 2017). The P300, which peaks at around 300 ms with an onset at around 250 ms after the stimulus, follows the processing of highly formulaic MWEs and also indexes reactions to predictable and unpredictable stimuli. For instance, the more positive P300s for words within idioms relative to other (literal) word combinations have been interpreted as associated with expectancies that arise during stimulus processing (Vespignani et al., 2010; Siyanova-Chanturia et al., 2017). Vespignani et al. (2010) interpreted the occurrence of the P300 as reflecting the match of the actual input (the idiom fragment) to the stored template (a specific configuration) retrieved from semantic

memory. Also Molinaro and Carreiras (2010), who studied the processing of Spanish collocations, interpreted the increased P300 as an index of initially recognizing that the sequence was a collocation, which lead to pre-activating the word completing the collocation. Finally, Siyanova-Chanturia et al. (2017) found the P300 component in the processing of binominals. They attributed the presence of the component to template matching. They explained that the component was detected, because when processing the binomial, participants expected particular unique words to occur, and they did not have to perform a lexical search for the item. The first part of the phrase was simply matched to a known template of this phrase retrieved from memory. The authors argued that the P300 and N400 components represented two different processing stages, one associated with the recognition of a unique, predictable and prefabricated routine (leading to the increased P300 amplitudes) and the other associated with facilitated processing and semantic integration (eliciting the reduced N400 amplitudes).

Overall, we have argued that MWEs (such as idioms, binominals, and collocations), which are characterized by a high degree of connectedness, are ubiquitous in native speaker discourse. MWEs can be identified in two ways in native speech and writing (Siyanova-Chanturia and Omidian, 2019): using native speaker judgments (the phraseological approach), and by applying measures based on word co-occurrence (the frequency-based approach). During comprehension tasks, native speakers process MWEs in their language faster than other less fixed expressions, which is revealed by their faster reaction times to such stimuli. MWE processing elicits specific ERP components. In particular, MWEs (such as idioms, binominals, and collocations) commonly evoke the reduced N400 component, meaning that they are highly predictable to native speakers. The processing of MWEs can also be accompanied by the P300 component, possibly indexing the matching of the input to the stored template. However, as claimed by Siyanova-Chanturia et al. (2017), the ERP research into MWEs is still in its infancy. Also, less is known about the comprehension of MWEs in speakers who are exposed to more than one language and there is a considerable shortage of ERP studies in this area.

## Cross-Linguistic Influences in the Use and Processing of Collocations

Speakers of two languages may not use and process MWEs similarly across their languages. In particular, they do not rely on MWEs in L2 speech and writing as much as in L1 (Pawley and Syder, 2014). Even advanced non-native L2 users produce fewer MWEs when compared to native speakers. However, some MWEs are over-represented, and others underrepresented in the L2 output (Durrant and Schmitt, 2009; Gozdawa-Gołębiowski and Opacki, 2018). This might depend on the L2 proficiency of the speakers. For instance, less proficient L2 users more eagerly rely on transparent phrasal verbs than on idioms and collocations (Kecskes, 2015), and may use L1 collocational patterns when speaking the L2 (Wray and Perkins, 2000). This suggests that the use of MWEs in the L2 is related to cross-linguistic influence (CLI) between the L1 and L2 of a particular speaker. CLI is

defined as “the influence of a person’s knowledge of one language on that persons’ knowledge or use of another language” (Jarvis and Pavlenko, 2008, p. 1).

Despite the linguistic evidence on the limited role of MWEs in L2 use and CLI from the L1, psycholinguistic evidence reveals that L2 users are sensitive to collocational patterns in both their L1 and L2. The processing of L2 collocations is faster than that of other L2 phrases, just like in the case of L1 collocations (Siyanova-Chanturia, 2017). However, the speed of processing L2 collocations depends on several factors. First of all, L2 users react faster to more frequent than to less frequent L2 collocations (Siyanova and Schmitt, 2008; Durrant and Doherty, 2010), which indicates that collocational processing might depend on the amount of exposure to L2 input. Second of all, incongruent L2 collocations (i.e., ones which do not have an equivalent in the L1) are more difficult to process than L2 collocations congruent with L1 patterns (Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013), and are also more difficult to learn (Szudarski and Conklin, 2014). As proposed, when the L2 user encounters an L2 collocation which is congruent with the L1 collocation, its appearance triggers the activation of the L1 translation equivalent and results in a faster recognition of the L2 collocation (Wolter and Gyllstad, 2011, 2013; Wolter and Yamashita, 2015). All these studies point to some carry over or CLI effects across the speakers’ languages in collocational processing and learning. Clearly, L1 knowledge influences the use and processing of the L2, indicating that both L1 and L2 systems are activated when L2 users comprehend MWEs.

However, assuming that both languages are indeed activated, a valid question about CLI, is to what extent the L2 collocational knowledge influences the knowledge and processing of collocations in the L1. Although the definition of CLI specifies that CLI effects can be bidirectional (see Cook, 2003; Cook, 2016; Jarvis and Pavlenko, 2008), there is surprisingly little evidence that collocational knowledge can be transferred not only from the speaker’s L1 to the L2, but also from the L2 to the L1. Such evidence has been studied mostly in migrant contexts and has been associated with language attrition – changes in the L1 system due to the exposure and use of the L2 (see Schmid and Köpke, 2017 for a discussion). For example, Marian and Kaushanskaya (2007) examined the patterns of lexical use and attrition in narratives produced by Russian migrants to the United States when they spoke English and Russian. The authors found examples of overt use of words borrowed from the language other than the one spoken and examples of CLI in the use of MWEs. Overall, bilinguals transferred more when speaking the L2 and borrowed more when speaking their L1. The only study on the CLI from L2 to L1 in the case of collocations that we are aware of is the one by Laufer (2003), who examined the acceptability of collocations transferred or calqued from Hebrew to Russian in the case of Russian migrants to Israel. In an acceptability judgment task, both Russian monolinguals and Russian-Hebrew bilingual migrants were asked to explicitly assess whether some sentences in Russian were acceptable or not. In Laufer (2003) study, many of the bilingual Russian migrants accepted collocational calques from their L2 Hebrew as correct in their L1 Russian. This was possibly due to the high exposure of

Russian speakers to the L2 Hebrew collocational patterns, which resulted in adopting the frequent ones within the L1.

However, Schmid and Köpke (2017) challenge the view that attrition or restructuring of the L1 under the influence of the L2 is a phenomenon specific only to migrant contexts. On the contrary, they claim that the process of CLI from L2 to L1 affects all bilinguals, and not only those who are immersed in using the L2 and make little use of their L1. Still, although there is a growing body of research on MWE processing and use among monolinguals, in the case of bilingual speakers most studies examined MWEs in their L2. Of these, studies of collocations are less numerous than those of other MWEs, and most research, especially such that relies on the use of ERP and eye-tracking, focuses on the processing of idioms (Siyanova-Chanturia, 2013, 2017). To our knowledge, there is nearly no research on the use of collocational patterns by L1 speakers living in the L1 environment who have a relatively high L2 proficiency. Our study aims to fill this gap.

## THE CURRENT STUDY

The goal of the study was to explore the sensitivity of Polish native speakers to CLI from their L2 English in the area of MWE and test whether the sensitivity is modulated by their L2 proficiency. Following Schmid and Köpke (2017), we assumed that due to contact with the L2, collocational patterns can be transferred from the L2 to the L1. We argue that this type of CLI is possible not only in the case of migrants immersed in the L2, but also in the case of L1 speakers. Although they live in the L1 environment, they have some knowledge of and contact with L2 English, a language of high social prestige and wide presence at the societal level (in the media, work environments, and education).

We assumed that as a result of intensive language contact (Winford, 2003) and CLI (Jarvis and Pavlenko, 2008), some MWEs, such as collocations can be borrowed and transferred from English to Polish. Here, we focused on collocational calques (loan translations) from L2 English, where the L2 words are replaced by semantically equivalent L1 Polish words (Haugen, 1950). Calques can be created by direct translation from the L2 to the L1. Polish collocational calques are then composed in accordance with the English pattern, but consist entirely of Polish words. Because calques contain only L1 words, they are less noticeable to Polish native speakers than foreign loanwords (Otwinowska-Kasztelanica, 2000), and can eventually become so frequent that they might get accepted as a part of the L1 language system and begin to be used by the speakers of the L1 according to the L2 pattern (Winford, 2003).

A good example of calques in Polish are novel verb-noun collocations based on English, e.g., *\*wziąć autobus* (“take the bus”), which have penetrated the Polish language and are sometimes used by Poles, even though they are incorrect from a prescriptive standpoint. In the case of *\*wziąć autobus*, a very similar Polish collocation exists, namely *wziąć taksówkę* (“take a taxi”), but it is restricted to only that particular type of transportation. We propose that the mechanism that leads to constructing such novel Polish collocational calques involves

replacing the original frequent L1 Polish verb (*pojechać* – “go”) by another frequent L1 verb (*wziąć* – “take”), typical of the English (L2) collocation (*\*wziąć autobus* – “take the bus”). This results in creating a meaningful word combination, superficially similar to the original Polish collocation.

In this study, we focused on verb-noun collocational calques from English (*\*wziąć autobus*) as compared with correct Polish collocations (*pojechać autobusem*) and with absurd expressions (*\*zjeść autobus* – “eat the bus”). We tested whether Polish L1 speakers with L2 English living in a Polish-speaking environment (1) judge the English collocational calques as acceptable, and (2) show sensitivity to the calques on a neural level as demonstrated by the ERPs. Moreover, we explored whether individual variability in L2 proficiency modulates the magnitude of these effects. To this end, we ran a behavioral study aiming to check the acceptability of the correct, calqued, and absurd collocations, and an ERP experiment testing the neural response to correct collocations vs. calques and absurd collocations. We expected that the collocational calques (*\*wziąć autobus*) and absurd expressions would be judged by bilingual native speakers of Polish as less acceptable than the correct collocations (*pojechać autobusem*). We also expected that the novel strings of words (calques and absurd expressions) would not be processed by bilingual native speakers of Polish similarly to the existing collocations (see Materials for the mean MI scores of the three groups of expressions and **Supplementary Appendix 1** for each MI score).

According to the ERP studies reviewed above, as participants read through a known (correct) collocation, the last word of the collocation should be predictable given the preceding words. It thus should lead to a reduced N400 component (Molinaro and Carreiras, 2010), compared to unknown or incorrect collocations. Also, relative to other non-fixed word combinations (not MWEs), the processing of correct collocations should activate the template matching mechanisms for linguistic information that is uniquely predictable, as revealed by the increased P300 component (Vespignani et al., 2010; Siyanova-Chanturia et al., 2017).

Following Molinaro and Carreiras (2010), we assumed that the effects would be time-locked to the last word of the collocation, which is the noun. In the correct Polish verb-noun collocations, on the presentation of the verb, the noun should become highly predictable and expected for native speakers of Polish. Thus, in our study, the correct Polish collocations (relative to English calques and absurd expressions) should evoke a reduced N400 time-locked to the onset of the noun, indicating facilitated semantic access or integration. The correct collocations should also evoke an increased P300 to the noun, indicating stronger activation of template matching than in the case of the non-fixed expressions (calques and absurd collocations). In contrast, the processing of the absurd collocations should evoke an increased N400 and a decreased P300 component, time-locked to the onset of the noun. The processing of the nouns in the calques should not be similar to those in the correct collocations demonstrating detection of anomaly, unless the L2 calques are already becoming accepted as a part of the L1 language system. We assume that the effects for collocational calques may be modulated by the participants' L2 English proficiency. More specifically, the higher

L2 proficiency, the less sensitivity to the calques should be observed (decreased N400 and increased P300 components).

## EXPERIMENT 1 – BEHAVIORAL ONLINE STUDY

### Methods

#### Participants

We recruited 35 native speakers of Polish, mean age 22.38 years (range 19–36;  $SD = 4.17$ ). Participants reported an average of almost 2 h (119 min) of English use daily ( $SD = 2.19$ ). Their self-rated proficiency in English equaled 5.12 (range 3–7;  $SD = 1.11$ ), as measured on a 7-point Likert scale (where 7 meant native-like proficiency). As for the proficiency range, 4 participants reported level 3, 6 participants reported level 4, 11 participants reported level 5, 11 participants reported level 6, and 3 participants reported level 7.

#### Materials

For the study, we created three lists of Polish language multi-word stimuli that consisted of verb-noun combinations. They included well-formed Polish collocations, collocational calques (verb-noun equivalents of well-formed English collocations translated word for word into Polish), and absurd expressions (verb-noun combinations where the verb did not match the noun semantically). To create the stimuli, we performed a frequency based cross-analysis of two national corpora, the British National Corpus (BNC, representing English) and the balanced subsections of the National Corpus of Polish (Narodowy Korpus Języka Polskiego, NKJP, representing Polish). For the BNC, we used the BYU search engine (Davies, 2004), annotated with the CLAWS tagset (Fligelstone et al., 1997), while for the NKJP, we used the Poliqarp search engine (Przepiórkowski et al., 2012) annotated with the Morfeusz tagset (Woliński, 2006). In both cases, we used regular expressions designed to locate verb-noun collocations that matched the specific query syntax of the given search engine (i.e., BNC vs. NKJP). The searches were lemmatic, meaning that they queried for all possible word forms. This was particularly important in the case of Polish, which has a very rich inflectional morphology. For both corpus analyses, the span of the search window relative to the potential collocate was set to six items and the searches used combinations of various word orders. These steps were taken to maximize the morpho-syntactic permutations that would be returned by our queries, e.g., the frequency and MI score (the strength of association between words; see section “What Are Multi-Word Expressions and How to Identify Them?”) for “give a presentation” would also be derived from “a presentation was given,” “giving a presentation in the morning,” “she gave several presentations,” etc.

Using this approach we created the three lists of stimuli taking four main steps. We first searched for English verb-noun collocations using a combination of two methods: the MI score (in the formulation of Stubbs, 1995) and n-grams, with false positives (e.g., machine annotation errors) vetted out by the trained linguists from within our team. We established which collocations are congruent and sufficiently represented in

English (i.e., we chose those with MI scores above 2.0 and high n-gram frequency in the BNC). The 1553 English collocations identified in this way were then translated into Polish word for word to create calques. Following this, we queried for the Polish equivalents of the collocational calques in the NKJP in order to verify that they were not typical in Polish (i.e., that they had a low MI score and n-gram frequency). This is how the initial list of calques was obtained.

Next, using the “phraseological approach” (i.e., native speaker intuitions, Siyanova-Chanturia and Omidian, 2019, see section “What Are Multi-Word Expressions and How to Identify Them?”), we picked 261 collocational calques that we evaluated as plausible in contemporary Polish (e.g., we heard some Polish people use them in conversations, or we considered them as highly likely to be used due to their semantic similarity to Polish MWE). We checked that they could be turned into their correct Polish equivalents by replacing the verb (e.g., *\*dać masaż*, “give a massage” – a calque vs. *robić masaż*, “make a massage” – a correct Polish collocation), resulting in a list of 183 phrases. Subsequently, we queried the NKJP to confirm high MI scores and n-gram frequency in the correct Polish collocations. To rule out co-incidental similarities between collocations in English and Polish, we also searched for analogs of the correct Polish collocations in the BNC and removed them from the list of well-formed Polish collocations. This is how the final list of 183 correct Polish collocations and collocational calques was created.

Then, to each pair of correct collocations and calques we added an absurd collocation by randomly assigning a verb to the noun (e.g., *\*sprzątnąć masaż*, “clean a massage”). To confirm that the combinations were semantically meaningless or paradoxical, we performed appropriate searches in the NKJP and BNC, verifying that the combinations had negative MI scores and no representative n-gram combinations. This is how the list of absurd collocations was created. Finally, we also controlled for the corpus frequency of the verbs used in the verb-noun expressions. As a result we ended up with three lists of expressions in Polish, each containing 183 items:

- (1) Correct, well-formed Polish verb-noun collocations (MI according to NKJP  $M = 9.01$ ,  $SD = 3.62$ ; verb frequency  $M = 3.41$ ,  $SD = 0.72$ ), e.g., *robić masaż*, “make a massage”; *mieć sens*, “have sense”; *pojechać autobusem*, “go by bus”;
- (2) Collocational calques from English, where the English verb was replaced by a Polish translation equivalent (MI according to NKJP  $M = 0.41$ ,  $SD = 1.96$ ; verb frequency  $M = 3.80$ ,  $SD = 0.88$ ), e.g., *\*dać masaż*, “give a massage”; *\*robić sens*, “make sense”; *\*wziąć autobus*, “take a bus”;
- (3) Absurd verb + noun expressions, where the verb did not collocate with the noun (verb frequency  $M = 3.31$ ,  $SD = 0.67$ ), e.g., *\*sprzątnąć masaż*, “clean a massage”; *\*zjadać sens*, “eat sense”; *\*kartkować autobus*, “browse the bus”.

Finally, each of the 183 collocations was embedded in a carrier sentence approximately 10 words long (following Molinaro and Carreiras, 2010). The same sentence was used across all three conditions (well-formed, calque and absurd) – which gave a

**TABLE 1** | Experiment 1: Descriptive statistics.

| Experiment 1 | Raw mean | SD   | Predicted mean |
|--------------|----------|------|----------------|
| Correct      | 4.58     | 0.73 | 4.58           |
| Calque       | 2.42     | 1.38 | 2.41           |
| Absurd       | 1.29     | 0.73 | 1.29           |

stimuli set consisting of 549 sentences (183 sentences \* 3 expression categories), for example:

- Fizjoterapeuta zrobił masaż/\*dał masaż/\*sprzątnął masaż kobiecie z uszkodzonym kręgosłupem.
- (The physiotherapist \*did a massage/gave a massage/\*cleaned a massage to a woman with a spinal injury).

All the phrases and the carrier sentences are presented in **Supplementary Appendix 1**. These sentences were divided into three lists, each containing 61 correct collocations, 61 calques and 61 absurd expressions, each presented in a different carrier sentence.

## Procedure

Experiment 1 was run as an online questionnaire in Google Forms. Participants were asked to judge whether the presented sentences were natural and acceptable in Polish. We used a 5-point Likert scale, where 1 meant “not natural at all” and 5 meant “perfectly natural in Polish.” Each participant assessed one randomly assigned sentence list, i.e., 183 sentences (containing 61 correct collocations, 61 calques and 61 absurd expressions).

## Statistical Analysis

In the current analysis we decided to account for some *a priori* predictions using contrasts (planned comparisons). Specifically, we established a repeated contrast matrix that allowed us to compare directly between the acceptability of the correct MWEs vs. absurd expressions, and the correct MWEs vs. calques. The analysis was performed using linear mixed-effects models, as implemented in the lme4 package (version 1.1.21; Bates et al., 2015) in R (R Core Team, 2017) using participants and sentences as crossed random effects. In the model, the dependent variable was the Acceptability score (from 1 to 5) for each sentence. As fixed effects we included: Type of collocation (correct, calque or absurd) and Proficiency (self-rated). To the categorical predictor of Type of collocation, we applied a repeated contrast such that we compared Correct vs. Calque, and Correct vs. Absurd.

The continuous predictor of Proficiency score was centered and scaled. We fitted the maximal model first (Barr et al., 2013) including the bobyqa optimizer, and in case of non-convergence or singularities we simplified it following recommendations outlines in Bates et al. (2018). The final model included the by-subject and by-sentence random intercept and correlated slope for Type of collocation. We considered as significant any fixed effect with an absolute *t*-value higher than 2.

## Results

The descriptive statistics for the behavioral study are presented in **Table 1**.

The analysis of the acceptability judgments showed a main effect of the comparison between Correct vs. Calques ( $t = -22.97$ ), such that the calques were assessed as less acceptable than the correct collocations. There was also a main effect of the comparison between Correct vs. Absurd ( $t = -57.76$ ), such that the absurd collocations were assessed as less acceptable than the correct ones (see **Table 2**). Importantly, no main effect of Proficiency or its interaction with collocation Type was observed.

## Discussion

In Experiment 1 we tested whether Polish bilingual native speakers living in a Polish-speaking environment would judge collocations calqued from English as acceptable. We also explored whether individual variability in L2 proficiency would modulate the judgments. Experiment 1 was a behavioral study aiming to check the acceptability of the correct collocations, calques, and absurd collocations. We expected that the collocational calques (\*wziąć autobus) and absurd expressions would be judged by native speakers of Polish as less acceptable than the correct collocations (pojechać autobusem). In accordance with our expectations, Polish native speakers found the native correct MWE fully acceptable (mean 4.59 on a 5 point Likert scale, where 5 meant “perfectly natural in Polish”). They also found absurd collocations unacceptable (mean 1.28, where 1 meant “not natural at all”). Collocational calques were closer to the mid-point of the scale (mean 2.39), and the difference between the correct MWEs and the calques was significant. Overall, this means that Polish native speakers judged the calques from English as significantly less acceptable than the correct MWEs. Importantly, participants’ self-rated proficiency in English did not have any effect on the judgments of any of the collocation types. Thus next, we conducted Experiment 2 to find out

**TABLE 2** | Experiment 1: Fixed and random effects for the LME model of the acceptability score.

|                                    | Estimate | Std. Error | t-value | By-participant SD | By-sentence SD |
|------------------------------------|----------|------------|---------|-------------------|----------------|
| Intercept                          | 4.58     | 0.05       | 86.06   | 0.21              | 0.47           |
| Correct vs. Calque                 | -2.17    | 0.09       | -22.67  | 0.38              | 0.87           |
| Correct vs. Absurd                 | -3.29    | 0.06       | -57.04  | 0.23              | 0.46           |
| Proficiency (self-reported)        | -0.01    | 0.04       | -0.01   | —                 | —              |
| (Correct vs. Calque) * Proficiency | -0.11    | 0.07       | -1.58   | —                 | —              |
| (Correct vs. Absurd) * Proficiency | -0.07    | 0.05       | -1.42   | —                 | —              |

whether there would be any differences in the processing of the three types of collocations by Polish speakers with the knowledge of English and whether their English proficiency would modulate the effects.

## EXPERIMENT 2 EEG STUDY

### Methods

#### Participants

Thirty new participants took part in the Experiment 2, all native speakers of Polish, aged 22–43 years ( $M = 25.47$ ,  $SD = 4.36$ ). They were recruited via a job-hunting internet portal and were paid for their participation. Their knowledge of English, ranged from intermediate to advanced (range 46.25–98.75;  $M = 70.86$ ,  $SD = 11.48$ ), as measured with LexTALE (Lemhöfer and Broersma, 2012). LexTALE scores are interpreted as follows: scores above 80 indicate an advanced to proficient L2 user (C1 and C2 level CEFR), scores in the range of 60–80 indicate an upper intermediate L2 user (B2 level) and scores below 59 indicate an intermediate level user (B1 level) and below (Lemhöfer and Broersma, 2012, p. 341). Thus, our participants' proficiency ranged between that of a pre-intermediate and proficient L2 user. Thus, our sample was characterized by quite a large variability in L2 English proficiency, between pre-intermediate and advanced, which actually adequately reflects the population of educated Poles.

#### Materials

From the 183 carrier sentences created for the behavioral study, we chose 120 (see **Supplementary Appendix 1**), for the ERP experiment. First, we eliminated sentences with 11 calques that were assessed as acceptable in Experiment 1 (the scaled acceptability rating for the sentence with the calque was over 0.78). We eliminated two sentences where the calqued expression was repeated due to technical error and 50 sentences where the collocational calques used the common verbs “take,” “adopt,” or “make,” to avoid over-representation of expressions with those words. The sentences were divided into three lists, each containing 40 well-formed collocations, 40 calques, and 40 absurd expressions. Within each list the carrier sentences did not repeat.

#### Procedure

Participants were tested individually in a dimly-lit, sound-attenuated room. They sat approximately 60–70 cm from the screen. They were asked to read the 120 sentences for comprehension one by one (i.e., one of the lists described above), which were visually presented on the screen word by word. The stimuli were presented centrally on a 17" CRT screen, in white letters on a dark-gray background. Each word appeared for 300 ms followed by a 300-ms blank screen and the sentence order was randomized. Participants had to answer Yes/No comprehension questions for 20% of the sentences containing the correct Polish collocation (see **Supplementary Appendix 1**). For example, following the sentence *Po imprezie musieliśmy posprzątać balagan i*

*wymienić szybę w kuchni* (After the party we had to clean up the mess and replace a glass pane in the kitchen) the question was *Czy impreza przebiegła spokojnie?* (Was the party peaceful?). Participants answered the questions by pressing the corresponding Yes/No button on the keyboard. The comprehension questions appeared randomly across the whole experiment.

#### Electrophysiological Recording

The electroencephalogram (EEG) was recorded at 256 Hz from 32 Ag/AgCl scalp electrodes positioned at the standard 10–20 locations, mounted in an elastic cap, using the Biosemi Active Two recording system. Electrodes were initially referenced online to the Common Mode Sense electrode located at the C1 electrode and re-referenced offline to the mean of the left and right mastoids. The horizontal and vertical electro-oculogram (EOG) was recorded bipolarly using electrodes placed below and above a participant's right eye and at the outer canthus of each eye, respectively.

The EEG signal was offline filtered with a band-pass filter (0.1–25 Hz frequency range; low cutoff slope: 24 dB/oct; high cutoff slope: 12 dB/oct). Before analyzing the data, artifacts (such as eye movements) were manually removed using independent component analyses (ICA) (Jung et al., 2000; Delorme et al., 2007). Additionally, other artifacts were defined as events in which there was a difference of  $\pm 100 \mu V$  in amplitude within less than 50 ms, or when the absolute amplitude exceeded  $\pm 100 \mu V$ . Trials with artifacts (2.99%) were rejected and recordings from electrodes with a high level of artifacts ( $> 50\%$ ) were interpolated using the average value of the group of nearest electrodes. We applied a baseline correction to the target nouns, creating epochs from  $-150$  to  $900$  ms. The accepted EEG epochs were obtained for each participant, each sentence, and each electrode across all conditions (Types of collocation: Correct, Calque, Absurd). Because the baseline derived directly from the noun onset would be affected by the processing of the previously presented verb, we assumed a baseline correction that was neutral with respect to the experimental manipulation. That is, we applied a baseline correction using the average EEG activity in the 150 ms prior to verb onset, which was pre-verbal. Importantly, because the baseline time window was defined as the 150 ms directly preceding the verb, we had to take into account any ERP effects that could have arisen before the noun was presented. Thus, our analyses encompassed the mean amplitude of both the verb and the noun. Consequently, we averaged and analyzed epochs with respect to the onset presentation of the verb of  $-150$  to  $900$  ms for the analysis of the verbs, and, because the noun was presented 500 after the onset of the verb, we extracted the epochs ( $-150-0$ )  $500-1400$  ms for the analysis of the nouns. That is, we extracted epochs from  $-150$  to  $900$  with respect to the presentation of the target words [ $-150-900$  in the case of verbs, and ( $-150-0$ )  $500-1400$  ms in the case of nouns]. Still, for the verbs in our collocations we did not have any specific predictions because the analysis at verb-level was purely exploratory.

Event-related potential extraction, averaging and cleaning were conducted with EEGLab (Delorme and Makeig, 2004) and ERPLab (Lopez-Calderon and Luck, 2014) MATLAB software toolboxes. We analyzed the P300 component for which we selected the 250–350 ms time window after word-onset on a cluster of centro-posterior electrodes: CP5, CP1, CP2, CP6, PO3, PZ, and PO4. We also analyzed the N400 component for which we selected the 350–450 ms time window after word-onset on a cluster of central electrodes: CZ, PZ, CP1, and CP2. The electrode clusters were chosen directly based on the literature on the N400 and P300 attesting to the scalp distribution of these components (Molinaro and Carreiras, 2010; Vespignani et al., 2010; Siyanova-Chanturia et al., 2017). We selected the Pz electrode for visualizing the effects because it was shared between the two clusters corresponding to each component (N400 and P300).

### Statistical Analysis

In the present study we explored whether calques would be processed differently from the correct L1 collocations. Still, to “situate” the calques we needed to evidence the difference between the correct and absurd conditions. In the analysis we again used contrasts (planned comparisons) to account for an *a priori* set of predictions (see section “The Current Study”). Specifically, we established a repeated contrast matrix that allowed us to compare directly between the processing of Correct vs. Absurd, and Correct vs. Calque. The reason to select those contrasts was theoretically driven. First, because in the literature on MWEs and collocations the P300 is used to show template matching mechanisms for linguistic information that is uniquely predictable, we assumed that the P300 should distinguish between both Correct and Absurd, and Correct and Calque. Also, since predictable word combinations should lead to a reduced N400 component compared to unknown or incorrect collocations, for the N400 the contrast between Correct vs. Absurd provided us with a “safety check” (predictable vs. unknown) and allowed to situate Calque relative to the Correct.

The analyses were performed using linear mixed effects models, as implemented in the lme4 package (version 1.1.21; Bates et al., 2015) in R (R Core Team, 2017) using participants and sentences as crossed random effects. In the models for electrophysiological data the dependent variable was the voltage (microvolts) for the selected time window and electrodes. As fixed effects for all the models we included: Type of collocation (correct, calque or absurd) and Proficiency (English LexTALE score). To the categorical predictor of Type of collocation we applied a treatment contrast where the correct condition was the baseline using the default contrast setting `contr.treatment` in R. Thus, we compared Correct vs. Calque, and Correct vs. Absurd. The continuous predictor of Proficiency was centered and scaled. We fitted the maximal model first (Barr et al., 2013) including the bobyqa optimizer. We considered as significant any fixed effect with an absolute *t*-value higher than 2.

## Results

### Verbs

The analysis of Verbs was purely exploratory. The descriptive statistics for Verbs in the P300 (250–350 ms), and N400 (350–450 ms) time-windows are presented in **Tables 3, 4**.

The final models for Verbs in the P300 (250–350 ms), and N400 (350–450 ms) time-windows included the by-subject random intercepts and correlated slopes for Type of collocation. They also included the by-sentence intercepts and the correlated slopes for Type of collocation, Proficiency, and the interaction between Type of collocation and Proficiency. For Verbs, the analysis revealed no significant effect or interaction in any of the time-windows (P300 and N400), and no interaction with L2 Proficiency, as presented in **Tables 5, 6**, respectively. The grand-averaged ERP waveforms for Verbs are illustrated in **Figure 1**.

### Nouns

The descriptive statistics for Nouns in the P300 (250–350 ms), and N400 (350–450 ms) time-windows are presented in **Tables 7, 8**.

Similarly to the Verbs, the final models for Nouns in the P300 and N400 time-windows included the by-subject random intercepts and correlated slopes for Type of collocation. They also included by-sentence intercepts and correlated slopes for Type of collocation, Proficiency, and the interaction between Type of collocation and Proficiency. The results for the respective time-windows (P300 and N400) will be discussed one by one below and presented in **Tables 9, 10**, respectively. The grand-averaged ERP waveforms for Nouns are illustrated in **Figure 1**.

#### The P300 effect (250–350 ms)

In the P300 region, the analysis revealed a main effect of comparison between Correct vs. Absurd ( $t = -1.99$ ) marginally significant, such that the Correct collocations yielded a more positive peak than the Absurd. No such effect was found between Correct and Calque, meaning that their waveforms were relatively similar (see **Table 4**, and **Figure 1**). The model did not find any main effect of Proficiency or its interaction with collocation Type.

#### The N400 effect (350–450 ms)

In this time window, the grand-averaged ERPs showed a distinct pattern based on Type of collocation. The

**TABLE 3** | Experiment 2: Descriptive statistics for the P300-Verbs.

|         | Raw mean | SD   | Predicted mean |
|---------|----------|------|----------------|
| Correct | 0.64     | 5.85 | 0.60           |
| Calque  | 0.80     | 5.71 | 0.83           |
| Absurd  | 0.61     | 5.74 | 0.57           |

**TABLE 4** | Experiment 2: Descriptive statistics for the N400-Verbs.

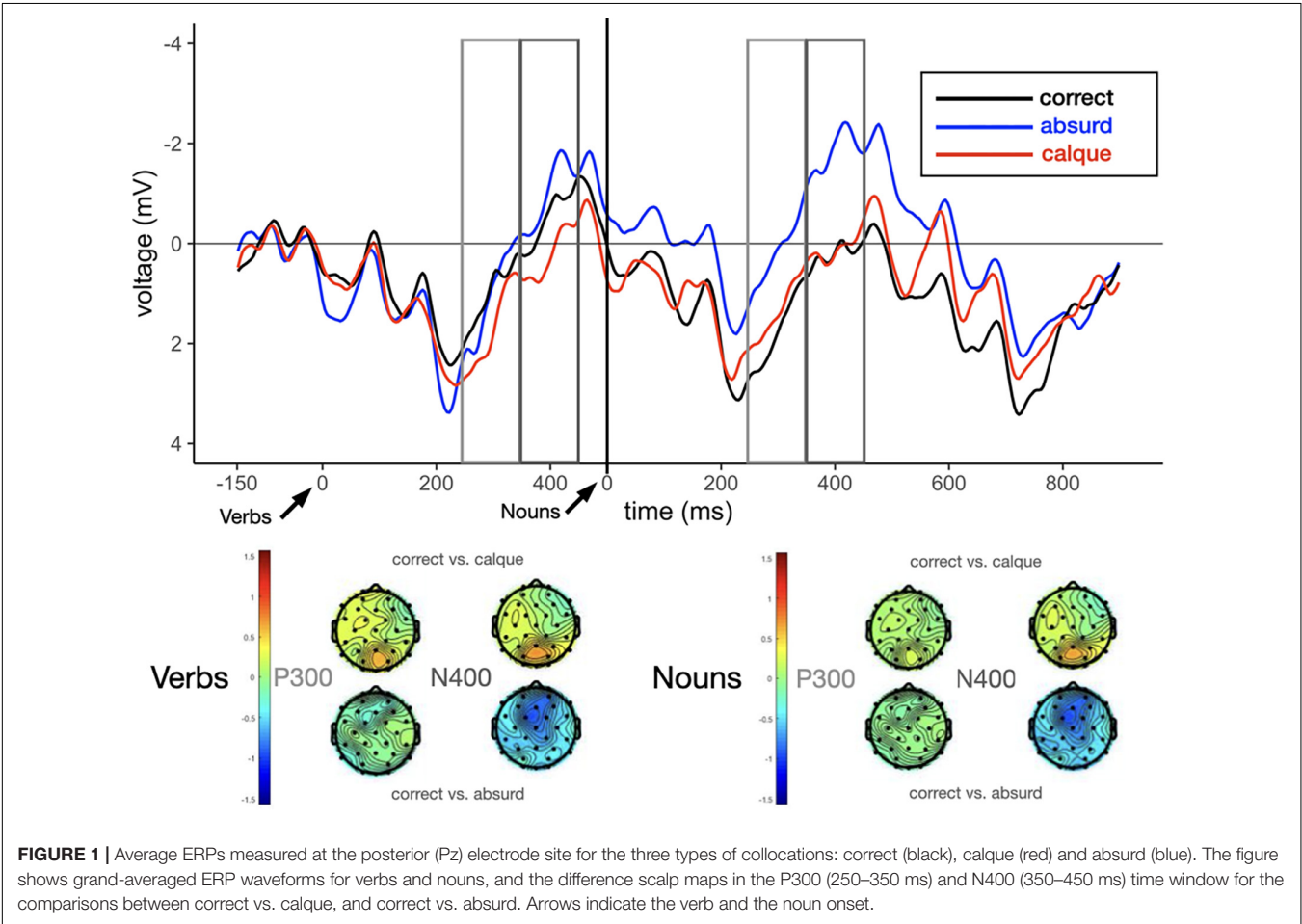
|         | Raw mean | SD   | Predicted mean |
|---------|----------|------|----------------|
| Correct | −0.36    | 6.78 | −0.38          |
| Calque  | −0.14    | 6.32 | −0.08          |
| Absurd  | −0.66    | 6.66 | −0.68          |

**TABLE 5 |** Experiment 2: Fixed and random effects for the LME model of P300-Verbs.

|                                    | Estimate | Std. Error | t-value | By-participant SD | By-sentence SD |
|------------------------------------|----------|------------|---------|-------------------|----------------|
| Intercept                          | 0.60     | 0.27       | 2.27    | 1.08              | 1.88           |
| Correct vs. Calque                 | 0.23     | 0.34       | 0.68    | 1.47              | 2.31           |
| Correct vs. Absurd                 | −0.03    | 0.26       | −0.10   | 0.80              | 2.33           |
| Proficiency (LexTALE)              | −0.00    | 0.25       | −0.00   | –                 | 1.68           |
| (Correct vs. Calque) * Proficiency | 0.08     | 0.35       | 0.24    | –                 | 2.39           |
| (Correct vs. Absurd) * Proficiency | 0.03     | 0.25       | 0.13    | –                 | 2.20           |

**TABLE 6 |** Experiment 2: Fixed and random effects for the LME model of N400-Verbs.

|                                    | Estimate | Std. Error | t-value | By-participant SD | By-sentence SD |
|------------------------------------|----------|------------|---------|-------------------|----------------|
| Intercept                          | −0.38    | 0.31       | −1.23   | 1.20              | 2.35           |
| Correct vs. Calque                 | 0.30     | 0.42       | 0.71    | 1.85              | 2.68           |
| Correct vs. Absurd                 | −0.30    | 0.40       | −0.76   | 1.42              | 3.18           |
| Proficiency (LexTALE)              | −0.30    | 0.29       | −1.07   | –                 | 1.92           |
| (Correct vs. Calque) * Proficiency | 0.36     | 0.42       | 0.86    | –                 | 2.64           |
| (Correct vs. Absurd) * Proficiency | 0.13     | 0.36       | 0.36    | –                 | 2.62           |



analysis revealed a main effect of the comparison between Correct and Absurd ( $t = -2.89$ ), such that for Correct collocations the negativity was reduced and for Absurd it was increased (see Table 10 and Figure 1). The

comparison between Correct and Calque was not significant, meaning that their waveforms were relatively similar. Again, no effect of Proficiency or its interaction with collocation Type was found.

**TABLE 7 |** Experiment 2: Descriptive statistics for the P300-Nouns.

|         | Raw mean | SD   | Predicted mean |
|---------|----------|------|----------------|
| Correct | 0.92     | 6.86 | 0.88           |
| Calque  | 0.62     | 7.16 | 0.64           |
| Absurd  | 0.17     | 7.09 | 0.06           |

**TABLE 8 |** Experiment 2: Descriptive statistics for the N400-Nouns.

|         | Raw mean | SD   | Predicted mean |
|---------|----------|------|----------------|
| Correct | 0.18     | 6.80 | 0.23           |
| Calque  | −0.33    | 7.03 | −0.30          |
| Absurd  | −0.77    | 7.25 | −0.75          |

## Discussion

In Experiment 2 we tested whether Polish bilingual native speakers living in a Polish-speaking environment would read collocational calques from English differently from correct Polish collocations. We also explored whether individual variability in L2 proficiency would modulate the effects. Experiment 2 was an ERP study aiming to compare the processing of the correct collocations, calques, and absurd collocations. We expected that the nouns in the collocational calques (*\*wziąć autobus*) and absurd expressions would be less predictable for native speakers of Polish than the nouns in the correct collocations (*pojechać autobusem*). In accordance with our expectations, reading the nouns in the correct Polish collocations yielded increased P300 and decreased N400 effects relative to the absurd collocations. However, contrary to our expectations, the collocational calques from English yielded results comparable to the correct MWEs. Similarly to Experiment 1, participants' proficiency in English did not modulate the processing of any of the collocation types.

## GENERAL DISCUSSION

This study focused on one type of multi-word expressions (MWEs), namely verb-noun collocations (e.g., *take a picture*). Its goal was to explore whether Polish native speakers are sensitive to cross-linguistic influence (CLI) from their L2 English in comprehending collocations and whether their sensitivity is modulated by their L2 proficiency. Although collocations are common in all natural languages, their processing is still under-researched in comparison to other MWEs (especially idioms, Siyanova-Chanturia, 2013).

Here, we zoomed in on the bilinguals' L1 rather than their L2, in contrast to many other studies on collocation processing by bilinguals and language learners, which examined native collocation patterns in the L2 use (e.g., Siyanova and Schmitt, 2008; Durrant and Doherty, 2010; Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013; Szudarski and Conklin, 2014). We did so to examine the possibility of reverse transfer, that is CLI from the L2 to the L1 (Jarvis and Pavlenko, 2008). CLI from the L2 to the L1 is ubiquitous in users of two languages (see Cook, 2003, Cook, 2016) and, according to Schmid and Köpke (2017), it is not restricted to contexts of migrants making little use of their L1 and immersed in the L2. However, reverse CLI from the L2 to L1 has not been previously explored in the case of MWEs, and especially collocations.

We investigated how the L1 verb-noun collocational patterns are processed by native speakers of Polish living in the L1 environment, assuming that they might be influenced by CLI from English, a frequent source of borrowing into Polish. We also assumed that the degree of CLI from English might be modulated by Polish speakers' level of English proficiency. To trace the reverse CLI from the L2 English in the L1 Polish, we asked whether English collocational calques are (1) acceptable to Poles and whether (2) they are

**TABLE 9 |** Experiment 2: Fixed and random effects for the LME model of P300-Nouns.

|                                    | Estimate | Std. Error | t-value | By-participant SD | By-sentence SD |
|------------------------------------|----------|------------|---------|-------------------|----------------|
| Intercept                          | 0.89     | 0.29       | 3.03    | 1.12              | 2.24           |
| Correct vs. Calque                 | −0.24    | 0.37       | −0.65   | 1.35              | 2.96           |
| Correct vs. Absurd                 | −0.82    | 0.41       | −1.99   | 1.58              | 3.15           |
| Proficiency (LexTALE)              | 0.20     | 0.28       | 0.72    | —                 | 2.11           |
| (Correct vs. Calque) * Proficiency | −0.01    | 0.40       | −0.02   | —                 | 3.34           |
| (Correct vs. Absurd) * Proficiency | 0.19     | 0.40       | 0.46    | —                 | 2.98           |

**TABLE 10 |** Experiment 2: Fixed and random effects for the LME model of N400-Nouns.

|                                    | Estimate | Std. Error | t-value | By-participant SD | By-sentence SD |
|------------------------------------|----------|------------|---------|-------------------|----------------|
| Intercept                          | 0.23     | 0.25       | 0.90    | 1.04              | 2.45           |
| Correct vs. Calque                 | −0.53    | 0.34       | −1.56   | 1.25              | 3.32           |
| Correct vs. Absurd                 | −0.98    | 0.41       | −2.40   | 1.75              | 3.67           |
| Proficiency (Lextale)              | 0.14     | 0.24       | 0.57    | —                 | 2.24           |
| (Correct vs. Calque) * Proficiency | 0.03     | 0.35       | 0.07    | —                 | 3.33           |
| (Correct vs. Absurd) * Proficiency | 0.25     | 0.41       | 0.62    | —                 | 3.54           |

processed similarly to Polish collocations by Polish speakers of varying English proficiency. To this end, we embedded three types of Polish verb-noun expressions into sentences: correct Polish collocations (e.g., *pojechać autobusem* – go by bus), collocational calques from English, where the verb was replaced by a Polish translation of the English verbs (e.g., *wziąć autobus* – take a bus) and absurd collocations (e.g., *zjeść autobus* – eat a bus). All collocations were identified using the frequency-based approach (MI score) and the phraseological approach (native speakers' intuitions, see Siyanova-Chanturia and Omidian, 2019).

To answer the first research question, in a behavioral study we presented L1 Polish participants of varying L2 English proficiency with the experimental sentences and asked them to rate the acceptability of those. The results were quite straightforward: a group of Polish native speakers explicitly judged most calqued expressions as less natural than the correct Polish collocations. The comparison of the sentence acceptability between the three types of stimuli revealed that the collocational calques were less acceptable than the correct Polish collocations, but more so than the sentences containing absurd expressions. The result for calques stands in contrast to Laufer (2003), whereby bilingual speakers accepted collocational calques from their L2 as correct in their L1. However, in Laufer's study the participants were immersed in the L2 environment, whereas those in our study were immersed in the L1. As such, our result suggests the collocational calques from English have not yet fully penetrated the Polish language, and Polish-English bilinguals living in the L1 environment are able to detect them as non-native in Polish. Interestingly, this effect was independent of participants' self-rated English proficiency, so, contrary to our expectations, even participants who knew English better, were still able to detect the calques from English.

To answer the second research question and explore the neural response to the calques from L2 to L1, we ran an ERP experiment in which participants were asked to read sentences that included calques, correct collocations and absurd expressions. On the basis of the behavioral study and previous research (Molinaro and Carreiras, 2010; Vespignani et al., 2010; Siyanova-Chanturia et al., 2017), we expected that native speakers of Polish would show sensitivity to the anomaly in both calques and absurd collocations. This sensitivity would be manifested by the modulations of N400 and P300. For the correct Polish collocations, we expected to obtain a reduced N400 to the nouns, which would reflect the fact that once a comprehender starts processing an MWE, the subsequent words of the MWE become more predictable. We also assumed a more positive amplitude of the P300 component to the nouns in the correct collocation condition, which would reflect a mechanism of template matching for MWEs. Conversely, for both calques and absurd collocations we expected a more negative N400, and a less pronounced P300 component to the nouns, relative to correct collocations. For the verbs we did not have any explicit predictions.

For the verbs, we did not find any significant effects of comparisons between the correct collocations and calques, as well as the correct and absurd collocations, time-locked to the P300

and N400 regions. All the effects that were found pertained to the nouns. However, contrary to our expectations, none of the observed effects to the nouns described below were modulated by the participants' L2 proficiency.

As expected, for the correct collocations, we indeed found an increased P300 amplitude for the correct collocations as compared with the absurd expressions and a reduced N400 to the nouns, which is in line the results from previous ERP research on MWE processing (Molinaro and Carreiras, 2010; Rommers et al., 2013; Vespignani et al., 2010; Siyanova-Chanturia et al., 2017). Following Siyanova-Chanturia et al. (2017), we interpreted the P300 for the correct collocations as typical of stronger template matching<sup>2</sup>. However, we do realize that the P300 results presented in our study are more supported by statistical models rather than by the waveforms (in comparison to the Pz electrode plots by Molinaro and Carreiras, 2010 or Siyanova-Chanturia et al., 2017). The reduced N400 amplitudes when processing correct Polish collocations (relative to the processing of calques and absurd collocations) provides evidence that native Polish collocations are processed with greater ease relative to other word combinations. In a given sentence context, the appearance of the verb belonging to an MWE made the noun that follows the verb in the MWE more predictable for Polish native speakers, possibly due to how the structure of collocations and their meaning are both stored in memory.

Although the ERP research on MWEs processing is still limited (Siyanova-Chanturia et al., 2017), the results presented here are partly in line with previous studies on MWEs, where the increased N400 was detected in response to novel, ill-formed, and metaphorical expressions (Molinaro and Carreiras, 2010; Vespignani et al., 2010). We found that absurd collocations elicited larger amplitudes of the N400 component than the correct collocations. However, in contrast to our expectations, we found there was a reduced amplitude of the N400 for nouns in the collocational calques, which mirrored the N400 to the correct collocations. In other words, the calques evoked similar neural responses to the correct collocations. This showed the participants' lack of sensitivity to the anomaly included in the calque and an effect of semantic integration of the nouns in both the correct collocations and calques. The effect might be explained in two ways.

First, it might mean that due to the borrowing and assimilation mechanisms (see Otwinowska-Kasztelanic, 2000; Winford, 2003) the English calques have already become part of the Polish language, and are common enough to be stored in memory and processed just like correct Polish collocations (so even the people who do not know English will treat and comprehend those calques on par with the correct Polish MWEs). However, this is not reflected at the behavioral level where sensitivity to the calques was still observed. Secondly, and more plausibly, the results mean that CLI from English plays a role, such

<sup>2</sup>It needs to be noted, however, that whether the P300 and N400 indeed reflect different underlying mechanism remains open for future research specifically targeted at this question. The template matching effectively implies the predictability of one word following another, and such predictability also lies at the core of the reduced N400. We thank reviewer 1 for bringing it up to our attention.

that Polish L2 users are sensitive to collocational patterns present in both their L1 and L2. The collocational calques, although still unacceptable in Polish, were congruent with the L2 English collocational patterns. The lack of significant differences between calques and the correct collocations at noun level in the P300 and N400 time windows might be a carry over effect of the congruency between the L2 collocational pattern and the calqued collocation in the L1. The effects of L1-L2 congruency were previously noted in studies on the processing of collocations in the L2 (e.g., Yamashita and Jiang, 2010; Wolter and Gyllstad, 2011, 2013; Szudarski and Conklin, 2014). Assuming that L1 knowledge influences the use and processing of the L2, the reverse pattern is also possible (Cook, 2003, Cook, 2016). The mechanism underlying the processing of the English calques in Polish might resemble the one proposed by Wolter and Gyllstad (2011, 2013). When the Polish user of English encountered a collocational calque from English, reading the verb in the collocation could trigger the activation of the collocation's English translation equivalent. This would lead to facilitated access to the L2 collocation. No difference in the P300 and N400 effects between the correct collocations and calques could, thus, result from participants' familiarity with calqued expression in their L2 English.

However, if calques are processed according to the L2 pattern due to CLI, as indicated by ERP data, then why do the behavioral results show speakers' sensitivity to them? The discrepancy in the results might be due to the very nature of the two tasks we used: acceptability judgments vs. reading for comprehension. Those two types of tasks involve different types of processing, the former being more explicit than the latter (Carrol and Conklin, 2020). In the behavioral study, participants were asked to identify the collocation, carry out a lexical search, and analyze the words to detect whether such a word combination is acceptable or not. Moreover, this task did not include any time limits, and most likely engaged participants' metalinguistic knowledge, thus increasing the chances of negative responses to calques that are "not quite right" in Polish. In contrast to the behavioral task, the ERP measurement did not rely on any acceptability decision and the task did not make participants explicitly focus on the collocations or consider their correctness in Polish. Participants had to read the sentences for comprehension (like in Molinaro and Carreiras, 2010; Vespignani et al., 2010; Rommers et al., 2013) and answer simple yes/no questions. In such a task, identifying the predicted word does not require active lexical search (for the argument see Vespignani et al., 2010). Thus, the system could accept the collocational calques as correct even though the verb did not quite match the native pattern. Still, in the case of nouns in the calques, enough semantic information was activated compared to the nouns in the correct collocations, which also resulted in reduced N400 components.

As a caveat, the system's acceptance of collocational calques as correct may have stemmed from participants' gradual exposure to the calques during the ERP experiment. In the **Supplementary Material** we present a series of **Additional Analyses** including

the trial number and the interaction of the condition with the trial number for the nouns. In a nutshell, all the models showed all the main effects obtained previously and there were no effects of the trial. However, the models also demonstrated interactions between the trial and condition. The results were most informative for the N400 to the noun, revealing a stable pattern for the correct collocations and the absurd collocations, but a gradual shift for the calques, which were first processed more like the absurd collocations, and only later like the correct ones. We can speculate that the results reflected some dynamics in the system, such that with longer exposure to the calques participants activated their L2 semantic networks and the memory traces for the L2 collocational patterns were becoming more active. Thus, the participants in our study, gradually exposed to the calques, were more likely to process them as the correct collocations.

Yet another possibility explaining the ERP results for the collocational calques is that the verbs in the calques were less semantically constraining than the verbs in the correct collocations, which meant that they connected to other words more readily and were plausible in a wider subset of sentences. Thus, possibly the presence of the calqued verbs in the sentence context was plausible enough to pre-activate a wide range of nouns to follow. This might depend on how constraining the sentence context was and how plausible a given verb was in that context. Unfortunately, we did not control for cloze-probability, which is a limitation of this research. As demonstrated by Carrol and Conklin (2020), both cloze-probability and MI score could indicate how easily a collocation can be integrated into the sentence. Also, future studies should test participants' collocational repertoire and consider it in the selection and creation of the materials.

## CONCLUSION

Due to the influence of the L2 on the L1, in some contexts, explicit judgments between native and non-native MWEs, and especially collocations, might become fuzzy (Laufer, 2003). Because evidence of this CLI phenomenon is scarce, we tested for the influence of L2 English, a prestigious global language, on the L1 Polish collocations of Poles living in an L1 environment. Although the behavioral results indicated that native speakers of Polish assessed the collocational calques from English as less acceptable than the well-formed Polish collocations, we did not find evidence that those two types of expressions differed considerably at early stages of processing (as evidenced by the P300 and N400 components), which preceded a conscious decision about the correctness of the sentences. This suggests a dissociation between explicit offline judgments and indices of online language processing measured by the ERPs. It might also indicate that the influence of English is so pervasive that Poles are becoming more and more oblivious to the collocational calques from this language. We believe that our study provides tentative evidence for this, and that calques from English have

penetrated the Polish language to the extent that they are no longer detected as anomalies, at least at the initial stages of processing. We conclude that the neural response to the calques and the explicit judgments about them indicates different levels of language processing that are to some extent dissociable. Crucially, the fact that the effects reported in the current study were not modulated by L2 proficiency suggests that borrowing and CLI from English, whose influence is ubiquitous, may lead to the assimilation of L2 calques in Polish. Calqued MWE might soon become fully integrated with the Polish language system, leading to language change at the societal level.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee at the Institute of Psychology, Jagiellonian University, approval number KE/10/032018 from 5.05.2018. The patients/participants provided their written informed consent to participate in this study.

## REFERENCES

- Arnon, I., and Snider, N. (2010). More than words: Frequency effects for multiword phrases. *J. Memory Lang.* 62, 67–82. doi: 10.1016/j.jml.2009.09.005
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Memory Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2018). Parsimonious mixed models. *arXiv* 1506.04967.
- Bates, D., Maechler, M., Bolker, B., Walker, S., and Haubo Bojesen Christensen, R. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7*. 2014.
- Carroll, G., and Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Lang. Speech* 63, 95–122.
- Cook, V. (2016). “Premises of multi-competence,” in *The Cambridge Handbook of Linguistic Multi-competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press).
- Cook, V. (ed.) (2003). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Davies, M. (2004). *British National Corpus*. Oxford: Oxford University Press.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004
- Durrant, P., and Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguist. Linguist. Theory* 6, 125–155.

## AUTHOR CONTRIBUTIONS

AO, MM, JD, JS, MO, and ZW: conceptualization. MM and MO: data curation. MM, AC, and MO: formal analysis. ZW: funding acquisition. MM: investigation. AO, MM, and JD: methodology. AO and MM: project administration and validation. MM, MO, and ZW: resources. MM and ZW: supervision. AC: visualization. AO, AC, and MO: writing – original draft. AO, MM, AC, JS, MO, and ZW: writing – review and editing. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was funded by a SONATA BIS National Science Centre grant awarded to ZW (2015/18/E/HS6/00428).

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of LangUsta lab members with experiment preparation and data collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.673761/full#supplementary-material>

- Durrant, P., and Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-Internat. Rev. Appl. Ling. Lang. Teach.* 47, 157–177. doi: 10.1515/iral.2009.007
- Fligelstone, S., Pacey, M., and Rayson, P. (1997). “How to generalize the task of annotation,” in *Corpus Annotation: Linguistic Information from Computer Text Corpora*, eds R. Garside, G. Leech, and A. McEnery (London: Longman), 122–136.
- Gozdawa-Golebiowski, R., and Opacki, M. (2018). Recurrent strings in corpus-based pedagogical research: A reappraisal of the field. *Glottodidactica. Internat. J. Appl. Ling.* 45, 134–149.
- Harper, C. (2008). *A Guide to Statistics: t-score and mutual information*. New York, NY: Harper Collins Publishers.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language* 26, 210–231.
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. London: Routledge.
- Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophys.* 111, 1745–1758. doi: 10.1016/s1388-2457(00)00386-2
- Kecskes, I. (2015). “Is the idiom principle plocked in bilingual L2 production?” in *Bilingual Figurative Language Processing*, eds R. Heredia and A. Cieśllicka (Cambridge: Cambridge University Press), 28–53.
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Kutas, M., and Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology* 11, 99–116. doi: 10.1016/0301-0511(80)90046-0
- Laufer, B. (2003). “The influence of L2 on L1 collocational knowledge and on L1 lexical diversity in written expression,” in *Effects of the Second Language on the First*, ed. V. Cook (Clevedon: Multilingual Matters), 19–31.

- Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0
- Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8:213. doi: 10.3389/fnhum.2014.00213
- Marian, V., and Kaushanskaya, M. (2007). Cross-linguistic transfer and borrowing in bilinguals. *Appl. Psycholinguist.* 28, 369–390.
- Molinaro, N., and Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biol. Psychol.* 83, 176–190. doi: 10.1016/j.biopsycho.2009.12.006
- Otwinowska-Kasztelanica, A. (2000). *A Study of the Lexico-semantic and Grammatical Influence of English on the Polish of the Younger Generation of Poles (19-35 years of age)*. Warszawa: Wydawnictwo Akademickie Dialog.
- Pawley, A., and Syder, F. H. (2014). *Two puzzles for linguistic theory: Nativelike selection and nativelike fluency*. In *Language and Communication*. Milton Park: Routledge, 203–239.
- Przepiórkowski, A., Bańko, M., Górski, R. L., Pęzik, P., Przepiórkowski, A., and Lewandowska-Tomaszczyk, B. (2012). *Narodowy Korpus Języka Polskiego*. Warsaw: Wydawnictwo Naukowe PWN.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing [Software]*. Vienna: R Foundation for Statistical Computing.
- Rommers, J., Dijkstra, T., and Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *J. Cogn. Neurosci.* 25, 762–776.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). “Multiword expressions: A pain in the neck for NLP,” in *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, (Berlin: Springer), 1–15.
- Schmid, M. S., and Köpcke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Ling. Appl. Biling.* 7, 637–667. doi: 10.1075/lab.17058.sch
- Seidlhofer, B. (2013). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Siyanova, A., and Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Can. Modern Lang. Rev.* 64, 429–458. doi: 10.3138/cmlr.64.3.429
- Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *Mental Lex.* 8, 245–268. doi: 10.1075/ml.8.2.06siy
- Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: a longitudinal study. *System* 53, 148–160. doi: 10.1016/j.system.2015.07.003
- Siyanova-Chanturia, A. (2017). Researching the teaching and learning of multi-word expressions. *Lang. Teach. Res.* 21, 289–297. doi: 10.1177/1362168817706842
- Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., and van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain Lang.* 175, 111–122. doi: 10.1016/j.bandl.2017.10.004
- Siyanova-Chanturia, A., and Omidian, T. (2019). “Key issues in researching multiword items,” in *The Routledge Handbook of Vocabulary Studies*, ed. S. Webb (London: Routledge), 529–544.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Funct. Lang.* 2, 23–55. doi: 10.1075/fol.2.1.03stu
- Szewczyk, J. M., and Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Lang. Cogn. Neurosci.* 33, 665–686. doi: 10.1080/23273798.2017.1401101
- Szudarski, P., and Conklin, K. (2014). Short-and long-term effects of rote rehearsal on ESL learners’ processing of L2 collocations. *TESOL Q.* 48, 833–842. doi: 10.1002/tesq.201
- Vespignani, F., Canal, P., Molinaro, N., and Fonda, S. (2010). Predictive mechanisms in idiom comprehension. *J. Cogn. Neurosci.* 22, 1682–1700. doi: 10.1162/jocn.2009.21293
- Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *J. Exp. Psychol.* 42:1632. doi: 10.1037/xlm0000259
- Vilkaitė, L., and Schmitt, N. (2019). Reading collocations in an L2: do collocation processing benefits extend to non-adjacent collocations? *Appl. Ling.* 40, 329–354. doi: 10.1093/applin/amx030
- Winford, D. (2003). *An Introduction to Contact Linguistics*. Oxford: Blackwell Publishing.
- Woliński, M. (2006). *Morfeusz—a practical tool for the morphological analysis of Polish in Intelligent information processing and web mining*. Berlin: Springer, 511–520.
- Wolter, B., and Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Appl. Linguist.* 32, 430–449.
- Wolter, B., and Gyllstad, H. (2013). Frequency of input and L2 collocational processing: a comparison of congruent and incongruent collocations. *Stud. Second Lang. Acquis.* 35, 451–482.
- Wolter, B., and Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Appl. Psycholing.* 36:1193. doi: 10.1017/S0142716414000113
- Wray, A. (2005). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state. *Ann. Rev. Appl. Ling.* 32, 231–254. doi: 10.1017/S026719051200013X
- Wray, A., and Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Lang. Comm.* 20, 1–28. doi: 10.1016/S0271-5309(99)00015-4
- Yamashita, J., and Jiang, N. A. N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Q.* 44, 647–668. doi: 10.5054/tq.2010.235998

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Otwinowska, Marecka, Casado, Durlík, Szewczyk, Opacki and Wodniecka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Crosslinguistic Influence (CLI) of Lexical Breadth and Depth in the Vocabulary of Bilingual Kindergarten Children – A Bilingual Intervention Study

Minna Lipner<sup>1</sup>, Sharon Armon-Lotem<sup>1,2\*</sup>, Joel Walters<sup>1</sup> and Carmit Altman<sup>2,3</sup>

<sup>1</sup> Department of English Literature and Linguistics, Bar-Ilan University, Ramat Gan, Israel, <sup>2</sup> Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat Gan, Israel, <sup>3</sup> School of Education, Bar-Ilan University, Ramat Gan, Israel

## OPEN ACCESS

### Edited by:

Tamar Degani,  
University of Haifa, Israel

### Reviewed by:

Amy Pace,  
University of Washington,  
United States  
Ishanti Gangopadhyay,  
Indiana University, United States

### \*Correspondence:

Sharon Armon-Lotem  
sharon.armon-lotem@biu.ac.il

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 24 February 2021

**Accepted:** 11 August 2021

**Published:** 30 September 2021

### Citation:

Lipner M, Armon-Lotem S,  
Walters J and Altman C (2021)  
Crosslinguistic Influence (CLI)  
of Lexical Breadth and Depth  
in the Vocabulary of Bilingual  
Kindergarten Children – A Bilingual  
Intervention Study.  
Front. Psychol. 12:671928.  
doi: 10.3389/fpsyg.2021.671928

**Introduction:** Research in recent years has explored the vocabulary size (lexical breadth) of bilingual children, but less is known about the richness of bilingual word knowledge (lexical depth), and about how knowledge of words in the two languages interact. This study explores how bilingual narrative intervention with vocabulary instruction in each language may modulate crosslinguistic influence (CLI) between the languages of bilingual kindergarten children, focusing on CLI of lexical knowledge, and which factors modulate performance.

**Methods:** Forty-one typically developing English-Hebrew bilingual children ( $M = 64.63$  months) participated. A bilingual adaptation of Story Champs narrative intervention program (Spencer and Petersen, 2012) was used to deliver vocabulary instruction in separate blocks of home language (HL) and school language (SL) sessions. Different intervention words were targeted in each language, but the children were tested on all target words in both languages. Lexical knowledge was assessed with a definition task four times throughout the study: prior to intervention, after each intervention block, and 4–6 weeks later. Learner characteristics (chronological age, age of onset of bilingualism and length of exposure) and proficiency in each language (standardized tests, familiarity with the vocabulary introduced in the intervention at baseline) were examined as possible modulators of performance.

**Results:** Children showed growth in lexical breadth and depth in their HL/English after HL intervention and in lexical breadth in the SL/Hebrew following SL intervention, with CLI for semantic depth observed via a qualitative analysis, but not quantitatively. Better HL/English performance was correlated with later AoB (and shorter SL exposure) and higher HL language proficiency scores. Children with higher HL/English proficiency responded better to the SL/Hebrew intervention, gaining more than those with lower English proficiency. Children with SL/Hebrew vocabulary dominance at the outset of the study also gained more from the HL/English intervention. No correlations were found between learner characteristics and SL performance.

**Discussion:** The current study indicates that bilingual narrative intervention with vocabulary instruction may be efficacious for improving the lexical breadth and depth of bilingual kindergarten children. It suggests that CLI may enhance bilingual children's language learning success, and points to the importance of strengthening both languages of bilingual children.

**Keywords:** crosslinguistic influence, lexical breadth, lexical depth, bilingualism, bilingual intervention, kindergarten children

## INTRODUCTION

Crosslinguistic influence (CLI), also known as crosslinguistic transfer, refers to the impact that a person's knowledge of a language has on knowledge or use of another language (Jarvis and Pavlenko, 2008; Serratrice, 2013), and it can be bidirectional. With regard to lexical CLI, there is evidence that the effects may be facilitative (e.g., Bilson et al., 2015), with words learned in one language providing conceptual and semantic bases upon which a translation equivalent may be learned. Bilingual children may transfer semantic information between languages (Foroodi-Nejad and Paradis, 2009; Goodrich et al., 2016) as they gradually learn translation equivalents (Pearson et al., 1999). Given the importance of CLI for language learning, this paper aims to gain a deeper understanding of its modulators.

Intervention studies have generally investigated the effects of intervention on lexical growth within one of a bilingual's languages (usually the societal language) (e.g., Carlo et al., 2004; Pham et al., 2018), but only a few have examined the effects of bilingual intervention (e.g., Restrepo et al., 2013; Armon-Lotem et al., 2020). In addition, most studies of vocabulary intervention have investigated effects on breadth of vocabulary (number of entries in a person's lexicon), without studying growth in lexical depth (the amount of information about a word). There is a critical need to understand both what is learned and how much is learned, given the importance of lexical depth for later reading comprehension (Bialystok, 2006; Ouellette, 2006; Proctor et al., 2012). In addition, the majority of research has been conducted with Spanish-English bilinguals in countries where English is the societal language. Few studies have been conducted in contexts where English is a high-status heritage language, i.e., where patterns of CLI could be different. Despite the high status of English in the present study, as a heritage language it may still be subject to attrition (Ardila et al., 2019). Moreover, even among bilinguals with typical language development, the two languages are usually not balanced (Pearson et al., 1993; Bernardini and Schlyter, 2004; Kupisch, 2008; Paradis, 2010; Hoff et al., 2012) and support for the home language is often limited. At the same time, the societal language is important, especially for academic success, and therefore also needs to be strengthened. Thus, intervention in preschool years targeting high level vocabulary in both languages may contribute to academic success and promote additive bilingualism, in which both languages may develop. Finally, studying children with typical language development provides the necessary baseline for future studies of children with atypical language development. Beyond the practical implications of this study, exploring the CLI of semantic knowledge as a result

of intervention is expected to contribute to understanding the organization of the lexical-semantic networks in this population.

The present paper explores how bilingual vocabulary intervention in the home and school languages (HL and SL) of bilingual kindergarteners impacts lexical breadth and depth in the HL and SL and may modulate the extent and nature of crosslinguistic transfer of semantic depth in the two languages. The aim is to document the intervention effects on CLI and explore the factors that may modulate these effects. In this vein, we examine children's learner characteristics such as bilingual experience (chronological age, age of onset of bilingualism) and language proficiency in each language as potential modulators of CLI.

## Lexical Breadth and Depth in the Vocabulary of Bilingual Kindergarten Children

In the bilingual lexicon, the specific words known in each language may be different (David and Wei, 2008; Bialystok et al., 2010). Some words may only be encountered at home, in one language, and others may only be used at school, in a different language (Hoff and Core, 2013). Exposure to input in two languages from different caregivers in different contexts may also contribute to the unique breadth of vocabulary found in bilinguals resulting from their exposure to a wide variety of input in the two languages. Hence, bilingual children might not have translation equivalents for every word in their lexicons, raising the question as to whether there is any transfer of lexical and semantic knowledge between the HL and SL. This question is discussed below in the section on CLI.

The bilingual lexicon can also be examined in terms of lexical breadth and depth (also known as semantic depth). Lexical breadth refers to the number of entries in one's lexicon. Lexical depth refers to the knowledge about those entries, viz. phonological, syntactic, collocational, morphological, semantic, pragmatic, and other information (Ordóñez et al., 2002). This information is acquired over time, through multiple encounters with each word. The relatively rich body of knowledge about vocabulary size, or lexical breadth, stands in contrast to the dearth of literature on lexical depth in bilinguals. Yet some scholars have noted that measuring lexical breadth without taking depth into consideration is of limited value (Wesche and Paribakht, 1996). At the initial stages of learning a word, children with "fast mapped" knowledge (Carey, 1978) know few aspects of a word, and may respond correctly on a receptive vocabulary measure but lack the ability to produce the word in conversation

or use this knowledge when reading (Hadley et al., 2016). Additional experience with individual words increases the depth of children's lexical knowledge. An information-rich context such as reading a book with pictures may enable gaining such additional knowledge from a single encounter (Hadley et al., 2016). The intervention in the current study presented words in a rich narrative context, with games and activities designed to broaden the contexts in which the words are found and deepen the knowledge of their use.

Tasks intended to tap into lexical depth include definitions (e.g., Snow, 1990), word association tasks (Schoonen and Verhallen, 2008; Sheng et al., 2012; Hai Weiss, 2020), and interviews that include giving definitions and answering questions about words (Ordóñez et al., 2002). Several studies of vocabulary depth in bilingual populations have found that depth is correlated across the two languages (Ordóñez et al., 2002; Dam et al., 2020), suggesting possible transfer of this kind of information. A recent study by Pham et al. (2018) investigated vocabulary learning following an HL intervention in several groups of bilingual children (ages 6–8). The study focused on several measures of vocabulary learning (identifying words in pictures, producing features of words, producing superordinate labels, and communicating word meaning effectively), which correspond to breadth and depth measures used in other studies (e.g., Ordóñez et al., 2002). Targets were eight words (four nouns, three verbs and one adjective), some of which were cognates. They found that the children improved in their HL on several measures of vocabulary, but only children with high HL proficiency who spoke a related language were able to transfer information to their SL, improving in definitions and communicating word meaning. These findings suggest that lexical depth may transfer.

However, with the exception of the Pham et al. (2018) study, there is a paucity of research investigating lexical depth and transfer effects in young bilingual children, especially following intervention. This gap motivated the present study in an effort to better understand the bilingual lexicon and explore ways to develop this important knowledge. It may lead to a better understanding of how bilingual children develop depth in their languages and how their language experiences and individual characteristics may contribute to this growth.

## Lexical and Semantic Crosslinguistic Influence (CLI)

Several models provide the theoretical basis for the current study. First, the Revised Hierarchical Model (Kroll and Stewart, 1994) posits shared conceptual representations for words as well as lexical links between words across languages. Indeed, much research has shown that when bilinguals use either one of their languages, words in both languages become active in parallel (Kroll and Ma, 2018). Accordingly, it may be posited that when a word is acquired in one language, knowledge of that word in the other language may be activated, facilitating transfer of semantic information. Second, Cummins' (1979, 1981) linguistic interdependence hypothesis and his further work (2000, 2008) maintain that bilinguals have a common

set of cognitive processes and a single representational system underlying their two languages, such that instruction in the HL should transfer to the SL and vice versa. Furthermore, he posits that knowing concepts in one language may expedite vocabulary learning in the second language, because the conceptual knowledge helps the child understand the meaning of the new unknown word. MacWhinney's Unified Competition Model (UCM; MacWhinney, 2005) maintains that when there is some kind of linguistic match between the HL and SL, language learners will attempt to transfer knowledge if it is close enough (even if it is not a complete match). Thus, positive transfer of lexical meaning in cases of words with the same or similar meanings is expected.

Work in past years has shown that words in a bilingual's two languages are mentally connected to each other, either directly (word-to-word) or through links to semantic representations (Jarvis and Pavlenko, 2008; Poulin-Dubois et al., 2013; DeAnda et al., 2016), thus providing the basis for transfer of lexical information between the two lexicons. MacSwan and Rolstad (2005) have suggested that young sequential bilinguals in the early stages of acquiring their second language may use the conceptual and lexical knowledge from their HL to assist them in learning words in SL. Indeed, Singh (2014) found crosslinguistic semantic priming effects from the dominant language to the non-dominant language in bilingual toddlers, indicating the existence of lexical stores that are conceptually related crosslinguistically.

A number of studies of bilinguals have shown that transfer of lexical knowledge occurs across languages and word learning in bilingual children is expedited by the second language (Ordóñez et al., 2002; Bilson et al., 2015; Goodrich et al., 2016; Armon-Lotem et al., 2020). However, most of these studies have focused exclusively on vocabulary breadth, without examining depth, leading to an incomplete understanding of lexical transfer.

## Modulators of CLI

In addition to documenting the presence of CLI, it is important to understand the circumstances in which it occurs. Research on the modulators of CLI has shed some light on this issue. One area of interest in research on CLI modulators is language dominance and proficiency. Bilingual children often differ in the language proficiency of their two languages, and in fact, most bilinguals are dominant in one or the other language (Pearson et al., 1993; Bernardini and Schlyter, 2004; Kupisch, 2008; Paradis, 2010; Hoff et al., 2012), which could potentially affect the extent and direction of CLI. According to the Revised Hierarchical Model (Kroll and Stewart, 1994; Kroll and Ma, 2018), the degree of bilingual fluency modulates the link between the lexicons and the shared conceptual representations of words, with more proficiency implying stronger lexical and conceptual links between concepts and words across languages, which will lead to enhanced performance on language tasks in the dominant language. Thus, according to this model, language dominance should play a role in determining cross-language performance, as vocabulary development in the SL should be expedited when concepts are already known in the HL.

However, despite the theoretical basis and evidence for shared semantic representations in the bilingual lexicon, studies on

vocabulary often focus on the factors leading to varying outcomes without relating to CLI (e.g., Paradis, 2011). One exception is a study by Altman et al. (2018), which examined the connection between vocabulary in one language and fast mapping in the other. In this study of Russian-Hebrew bilingual kindergarteners (ages 5–6), the authors found that the Russian (HL) dominant children used receptive vocabulary in Russian to support fast mapping in SL/Hebrew. A study by Kan and Kohnert (2012) also found crosslinguistic relations between HL vocabulary and word learning ability in SL in young sequential bilinguals (ages 3–5), with larger HL vocabularies supporting SL learning. However, the study did not investigate the factors affecting variation in CLI. The authors speculated that the children's home environments may have accounted for the differences.

Another factor impacting CLI is age of onset of bilingualism (AoB), which is often related to language proficiency and dominance. Meir et al. (2017) found evidence for bidirectional CLI in the morphosyntax of 5–6 year-old children, showing that participants with earlier AoB to SL (Hebrew) (0–23 months) mastered the contrastive structures (articles, perfect aspect, case inflections) better in SL and performed worse in HL (Russian), while those with later AoB had better performance in HL and weaker acquisition of the same structures in SL. Thus, AoB affected the degree of CLI, with more impact of the HL occurring with later AoB for structures that differed in the two languages. Likewise, Chondrogianni and Marinis (2011) evaluated the influence of AoB, length of exposure, and other factors on children's acquisition of SL vocabulary (and morphosyntax) and found that age of onset and length of exposure were significant predictors of vocabulary level: later AoB and longer exposure led to enhanced performance, suggesting that increased exposure may facilitate vocabulary development. In contrast, Unsworth (2016) did not find effects of AoB in the SL acquisition of vocabulary when she compared bilinguals with AoB of 1–3 years versus 4–7 years. These mixed findings for AoB could be related to the relation between quantity of input, vocabulary knowledge, and transfer effects (Pearson et al., 1997; Vermeer, 2001; Poulin-Dubois et al., 2013), which is beyond the scope of the present study.

Some intervention studies have focused on the language of instruction as a variable in performance (Ebert et al., 2014). Lugo-Neris et al. (2015) investigated the order of the language of intervention for a group of Spanish-English bilinguals at risk for language impairment, and found that those who received initial intervention in HL/Spanish (the more proficient language for most of the children) made larger gains than those who began in SL/English. This finding provided the motivation in the current study for intervening first in the HL/English, as it was the more dominant language for most of the children. To sum up, research on factors modulating CLI has not produced a clear picture of the conditions that facilitate crosslinguistic transfer of lexical knowledge.

## Vocabulary Intervention Studies

In addition to the intervention studies mentioned above, which focused on language of instruction as a variable, a number of other vocabulary intervention studies have been conducted

among bilingual children. Carlo et al. (2004) conducted a vocabulary intervention with monolingual and bilingual fifth graders, providing the bilingual children with materials in their native language (Spanish) to preview before the English language intervention. They found that all children in the intervention group improved on vocabulary measures including measures of vocabulary depth, as well as on comprehension, with gains for bilinguals similar to those of the monolingual students. Investigating the role of the language of vocabulary instruction with preschool Spanish-English children, Mendez et al. (2015) compared the effects of instruction in English only versus bilingual instruction using various instructional strategies. They found that immediately following instruction, the children in the bilingual instructional condition knew significantly more vocabulary words in SL than the children in the English-only group. The study also reported, as did previous research (Rolstad et al., 2005; Restrepo et al., 2013), that children exposed to bilingual instruction gained vocabulary in HL as well. The authors speculated that use of HL for presenting words before introducing them in SL allowed children to use their existing linguistic and conceptual HL knowledge to facilitate SL acquisition.

However, these studies introduced the same words in both languages, thereby drawing explicit connections between the words. The question remains as to whether words and concepts introduced in one language only will be connected to information about the words in the other language without explicit connections being made between them. That is, if a concept is introduced in one language, according to the Revised Hierarchical Model (Kroll and Stewart, 1994), the lexical knowledge of that concept should be activated (Kroll and Ma, 2018), allowing for transfer of the new knowledge. The current study allows for exploration of this hypothesis.

A recent study by Armon-Lotem et al. (2020) provides some evidence to support this assumption. Sixteen English-Hebrew bilingual preschoolers underwent a bilingual narrative intervention (BINARI) with vocabulary instruction, using a design similar to the one reported in the present study. The children made progress in vocabulary breadth with gains in the language of intervention as expected. Cross-linguistic gains were observed in SL/Hebrew following the HL/English intervention, but no gains were observed in HL/English following the SL/Hebrew intervention. This study was conducted with a small group of children with no control group and did not distinguish between words taught in the intervention and translation equivalents. Further work is needed to determine whether and under what circumstances information may transfer in both directions as a way of better understanding the nature of vocabulary depth in bilingual children. The Pham et al. (2018) study reviewed above provides some evidence of transfer of lexical information, including depth, but it was conducted with a small number of words, half of which were cognates.

Taken together, these studies demonstrate that vocabulary interventions can be successfully implemented with bilingual populations. Intervention studies focusing primarily on vocabulary indicate that bilingual intervention has positive effects on both languages, whereas intervention only in SL

impacts the SL only and may have detrimental effects on HL (Restrepo et al., 2013). Armon-Lotem et al. (2020) found that all children showed gains in SL following intervention in HL, but only children who were relatively stronger in HL gained in HL following SL intervention. This result is one of the motivations for intervening in and testing both languages. The impact of intervention in each language separately and the cumulative effect of intervention in two languages will be addressed to better understand the nature of transfer of lexical knowledge across languages.

## The Present Study

The present study investigated the effects of a bilingual narrative intervention with embedded vocabulary instruction on the lexical breadth and depth of English-Hebrew bilingual kindergarten children. Administering intervention in both languages for different words enabled an examination of CLI from one language to the other, as all the words were tested in both languages. Progress was measured with a definitions task, which was scored to measure growth in both vocabulary breadth and depth. The following questions are addressed:

1. Does bilingual narrative intervention (BINARI), in particular, vocabulary instruction, improve lexical breadth and depth in the HL and SL of English-Hebrew preschool children?
2. Do the learner characteristics of language dominance, higher proficiency, and AoB predict CLI of lexical breadth and depth?
3. How is CLI from the HL to the SL and from the SL to the HL manifested in lexical breadth and depth?

BINARI is anticipated to enhance lexical knowledge in children in both languages by increasing the breadth and depth of their vocabulary. CLI is expected to be greater for transfer of lexical information (depth) for familiar concepts, since increasing depth for an already known lexical form may occur more readily than for a novel form, as seen in previous research (Pham et al., 2018). By contrast, increasing lexical breadth with new concepts introduced in one language might take longer to impact the other language and lead to acquiring a word in the that language without further support. Finally, language dominance, higher proficiency and AoB (as a correlate of length of exposure) are predicted to facilitate CLI.

## MATERIALS AND METHODS

### Participants

Forty-one typically developing (TLD) English-Hebrew bilingual preschool children aged 5–6 years old ( $M = 64.63$  months) participated in the study. The children were recruited from preschools located in areas with a concentration of English-Hebrew bilingual families. Consent forms were sent to the parents of approximately 200 children in six kindergartens. Forty-three parents gave consent for their children to participate in the study and filled in a parental questionnaire about the children's demographic and linguistic background. Two children were excluded due to very low performance on both the English and

Hebrew screening test, which might indicate atypical language development (Håkansson et al., 2003). The remaining forty-one children were randomly assigned to two groups, one experimental, whose members received BINARI, the other, a control group. Criteria for inclusion in the study were: (1) one of the parents had to be a native English speaker; (2) the child had to score at or above local bilingual standards (see below) on either the English or Hebrew standardized language screening test (Armon-Lotem and Meir, 2016; Armon-Lotem et al., 2021); (3) the child had to score above 85 on the Raven Progressive Matrices non-verbal intelligence test (Raven et al., 1998); (4) the child had to be exposed to the L2 for at least 24 months; and (5) the child did not have any history of a hearing impairment or parental concerns about language. Written informed consent was obtained from the parents, who also provided information about their child's language development and background. Children expressed oral assent to participate in the study and were allowed to terminate participation at any time. The study was approved by the Bar-Ilan University IRB and the office of the Chief Scientist of the Israeli Ministry of Education.

In order to assess the language performance of bilingual children in their home language (HL/English), the Clinical Evaluation of Language Fundamentals (CELF Preschool-2; Wiig et al., 2004) was used. The CELF-Preschool-2 (Wiig et al., 2004) consists of six subtests for concepts and following directions, word structure, expressive vocabulary, recalling sentences, sentence structure, and receptive and expressive word classes. Local bilingual standards for the Core Language Scores (CLS) of the CELF-Preschool-2 (with a cutoff point of  $-1.25$  SD) are available for English-Hebrew bilinguals (Rose, 2018; Armon-Lotem et al., 2021). Data from 240 typically developing English-Hebrew bilingual children aged 5;0–6;5 years have been used to set the local standards for the CLS taking into account chronological age and age of onset of bilingualism, identifying the means and SDs which indicate typical development in English as a HL in Israel. To assess the children's language performance in the school language (SL)/Hebrew, the *Goralnik Screening Test for Hebrew* (Goralnik, 1995) was administered. The test includes six subtests: sentence repetition, comprehension, expression, pronunciation, vocabulary, and storytelling. The scoring totals 180 points, 30 for each subtest. Local bilingual standards for the Goralnik raw score were used with a cutoff point of  $-1.25$  SD (Iluz-Cohen and Armon-Lotem, 2013; Armon-Lotem, 2014; Altman et al., 2016). Data from 443 bilingual children aged 5;0–6;5 years speaking Hebrew as SL have been used to set the local standards for the Goralnik, taking into account chronological age and age of onset of bilingualism, by identifying the means and SDs which indicate typical development for SL/Hebrew in Israel.

**Table 1** displays the children's mean ages at the onset of the study, their age of onset of bilingualism (AoB), i.e., when they were first exposed to SL/Hebrew, years of mothers' education as a proxy for socioeconomic status, and their results on the standardized language tests (mean Core Language Scores [CLS] for the CELF on a standard scale where the mean is 100 and SD is 15, and the mean raw scores for the Goralnik). The information from the standardized testing is used to compare the performance of the experimental group and the control group, not to compare

**TABLE 1** | Background information for participants.

| Group                         | Mean age<br>(in<br>months) | Mean AoB<br>(in<br>months) | Mean years of<br>mothers'<br>education | Raven<br>standard<br>score     | HL/English<br>proficiency:<br>CELF–CLS | SL/Hebrew<br>proficiency Goralnik<br>raw scores |
|-------------------------------|----------------------------|----------------------------|----------------------------------------|--------------------------------|----------------------------------------|-------------------------------------------------|
| Experimental ( <i>n</i> = 22) | 64.82<br>(3.74)<br>(58–72) | 17.18<br>(15.64)<br>(0–38) | 16.86 (2.44)<br>(12–21)                | 116.89<br>(12.96)<br>(90–132)  | 97.77 (11.96)<br>(81–114)              | 129.14 (17.65)<br>(90–157)                      |
| Control ( <i>n</i> = 19)      | 64.42<br>(5.34)<br>(57–75) | 26.84<br>(10.86)<br>(4–40) | 16.05 (2.55)<br>(12–21)                | 120.46<br>(11.70)<br>(100–138) | 99.74 (10.78)<br>(81–125)              | 125.58 (13.14)<br>(106–143)                     |
| Totals ( <i>n</i> = 41)       | 64.63<br>(4.49)<br>(57–75) | 21.66<br>(14.33)<br>(0–40) | 16.49 (2.49)<br>(12–21)                | 118.34<br>(12.40)<br>(90–138)  | 96.68 (11.33)<br>(79–125)              | 127.49 (15.63)<br>(90–157)                      |

HL, home language/English; SL, school language/Hebrew; AoB, age of onset of bilingualism; HL assessed via CELF (Wiig et al., 2004); SL proficiency assessed via Goralnik (1995). Raven standard score is based on a subset of 17 children in the experimental group and 13 in the control group as no norms were found for the younger children.

the two languages within each group, as the two measures are not comparable. The standardized tests make it possible to compare the children's performance in each language within the group to assess the impact of proficiency level on benefits from the intervention. No significant differences were found between the experimental and control groups in age and years of mothers' education. The age of onset of bilingualism (AoB) differed significantly between the two groups, with the experimental group having a lower AoB than the controls. However, there was no significant difference between the groups in their language proficiency in either language.

## Materials and Procedure

### Overview

The BINARI intervention utilized in the study was based on *Puente de Cuentos* (Spencer et al., 2017), a Spanish-English bilingual adaptation of the *Story Champs* program (Spencer and Petersen, 2012), designed to enhance the narrative and vocabulary skills of Spanish-English bilingual preschool children (for details see Petersen and Spencer, 2016). Several adaptations were made to the program for use in the present study. First, the most culturally appropriate stories were selected, translated and adapted linguistically and culturally to the Israeli context. Second, ten stories were translated into Hebrew [six for the intervention sessions, and four for progress monitoring (PM)]. Third, the original five-picture format of the stories was expanded to include six pictures in order to better conform to the story grammar literature. Finally, following the results of a pilot study with 16 children (Armon-Lotem et al., 2020), six of the twenty-four vocabulary items were changed to better reflect the culture and level of this population.

The target vocabulary words were chosen based on the original *Puente de Cuentos* intervention, with adaptations made based on the pilot study. Six verbs and six adjectives were selected for intervention in each language to cover a range of difficulty (based on the pilot study and in consultation with preschool teachers). For the Hebrew adaptation, bilingualism experts and experienced preschool teachers were consulted to choose words that would be of an appropriate level in the absence of Hebrew frequency lists. The intent was to leave room for improvement on the one hand,

but to make sure the task was not too frustrating on the other. This variation in difficulty also allowed for the diversity in the children's abilities. In general, verbs and adjectives are considered more challenging for this age group (Johnson and Anglin, 1995; Blackwell, 2005) and are thus considered appropriate targets for language enrichment.

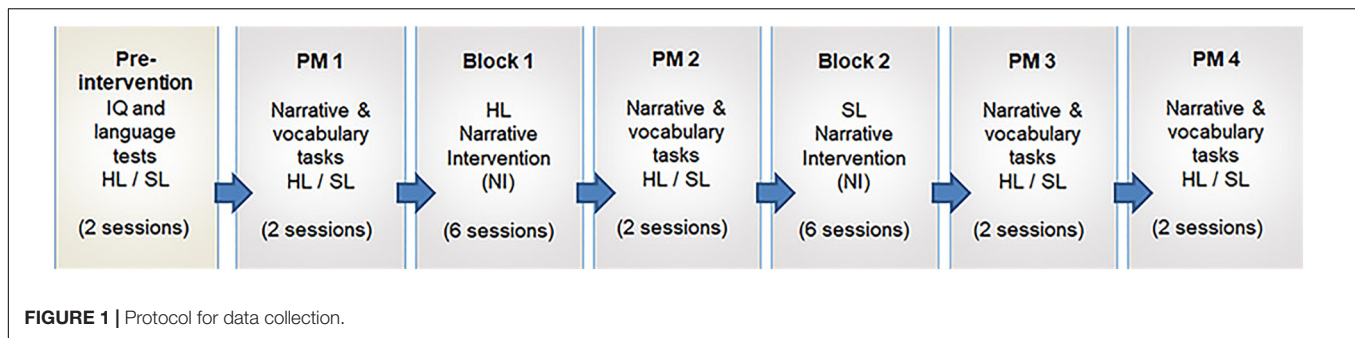
### Design

The intervention consisted of two blocks of six group sessions each: the first block was conducted in the HL (English), and the second in the SL (Hebrew). The children were evaluated four times in progress monitoring (PM) sessions: before the intervention, after each block of intervention, and 4–6 weeks after the end of the intervention, to check for maintenance. PM sessions took place in the children's preschools or online via Zoom® in their homes (see "Covid-19 adaptations" for details below). The sessions were audio recorded. All research assistants were native speakers of the target languages (two for each language) and had extensive experience working with children this age. **Figure 1** summarizes the protocol for data collection. The control group followed the same PM sessions protocol as the experimental group and at similar intervals, but did not undergo any intervention.

The majority (95%) of the intervention sessions took place in the children's preschools, one to three times a week, depending on the preschool schedule, and totaled six sessions over 2–3 weeks in each language. Each session lasted 20–25 min.

### Procedure

Intervention was conducted in groups consisting of three-four children. Each session began with a short introduction about the purpose of the session and a very brief review of the words learned in the previous session in that language. Then a new story was introduced, and the accompanying activities from the *Puente de Cuentos* manual (Spencer et al., 2017) were carried out as well as a supplementary vocabulary activity for each word at the end of the session. The targeted vocabulary words were first introduced in the context of a story, and then practiced individually, first with reference to the story content, and then in other contexts, thereby enriching the children's knowledge of the words. Each session included multiple repetitions of the story, with telling and



retelling by the experimenter and children. Thus, the words were introduced in a context rich in information and visual stimuli (Hadley et al., 2016) designed to develop vocabulary depth as well as enhance breadth. See **Supplementary Appendix A** for a sample intervention protocol.

### Vocabulary Knowledge Task

The PM sessions included a Vocabulary Knowledge Task (designed for the present study and used in the pilot) to test the children's knowledge of the target words and their translation equivalents. The vocabulary task was developed to assess the semantic content of the children's definitions (Korat et al., 2014), as opposed to definitional form (Benelli et al., 2006).

The task was designed to parallel the intervention. In the intervention, the same 24 target words were taught to all participants, 12 in the English intervention and 12 in the Hebrew intervention. All 24 words were tested in the vocabulary task in each language: 12 English target words and 12 English translation equivalents for the Hebrew target words were used in English testing and 12 Hebrew target words and 12 Hebrew translation equivalents of the English target words were used in Hebrew testing. This procedure was designed to allow assessment of language transfer effects in both directions. The only difference between the Hebrew and English tasks were the stimulus sentences, which contained different content to ensure that the children would focus on the target words and not on the sentence contexts. **Supplementary Appendix B** contains the Vocabulary Knowledge Task for English and Hebrew.

In the task children were asked to define the target words and the translation equivalents (verbs and adjectives) introduced in the BINARI sessions. Generally, a child's ability to explicitly formulate a qualitative definition is considered a clear indication that the word is known (Johnson and Anglin, 1995). Definition tasks have been used to measure depth of vocabulary knowledge (e.g., Read, 1993; Verhallen and Schoonen, 1993; Ordóñez et al., 2002). In the task, the words were read aloud to the children in short sentences designed to provide minimal contextual clues. The experimenter began by saying, "We're going to play a game with words and sentences. I'm going to read you sentences with words. I want you to tell me what the words mean – whichever ones you know." Children practiced on a sample word and sentence before starting the task. After each sentence, the experimenter asked what the target word meant. For example, for the target word "scrubbed," the prompt was: "The boy *scrubbed*

the floor. What does *scrubbed* mean?" Experimenters did not comment on responses. However, if the response was ambiguous, or if the child used a gesture instead of defining it verbally, the experimenter encouraged the child to elaborate by asking the child to explain the item verbally. The test was administered to all participants in the same order. Each language was tested on a different day by native speakers of that language. Responses to the vocabulary task were recorded manually during the sessions, with audio recordings used as backup.

### Adaptations for Covid-19

Overall, 95% of the intervention was completed in face-to-face sessions. Following the onset of Covid-19, kindergartens were closed, so the final one or two sessions of the intervention were completed individually online for some of the children ( $N = 12$ ). In addition, PM3 and PM4 were administered online for the experimental group, and PM2-4 were conducted online for the control group.

The online versions of the final intervention sessions and PMs were created to be as similar as possible to the face-to-face version and were discussed amongst experimenters for consistency and practiced before data collection. At the end of the first online session, each child was asked how he/she felt about doing it online. All children had experienced online learning from the beginning of the first lockdown, and most reported that they enjoyed it (some even said they liked it more than in the kindergarten). The four research assistants reported that the children cooperated very well online, and that they did not notice differences in their performance (other than the expected improvement following intervention).

### Data Scoring and Analysis

#### Coding and Scoring

Children's responses were rated for their expressed knowledge of the words. The responses were coded as full definition, partial definition, codeswitching, gesture, wrong definition and "I don't know." A response was counted as a codeswitch when a codeswitched element was a meaningful and relevant part of the definition; all codeswitches consisted of single words. For example, when asked to define "heavy" in English, a child said "*kaved*" (heavy in Hebrew). Expressive lexical knowledge includes all answers that show knowledge of the word as well as codeswitching. Scoring was based on the content of the child's response, not on the form of the definition. Scores for

breadth were 0–1, reflecting no knowledge versus some/complete knowledge including codeswitching. Scores for depth were 0–1–2, where no response or an error received a score of zero (0), a partially correct answer (including associations and incomplete explanations) received one point (1), and complete knowledge received a score of two (2). A score of two was given only to responses that included a precise and/or complete definition and was perceived as an indicator of greater depth of knowledge. Less complete or precise responses that nevertheless showed some knowledge of a word received one point. Thus, a score of one on the breadth scale would receive a score of one or two on the depth scale. Codeswitches were scored as 0 for lexical depth as they were not indicative of expressive vocabulary in the tested language. Gestures were scored as 0 for both breadth and depth as they indicate non-verbal knowledge.

**Tables 2A,B** outline the coding and scoring for the lexical breadth (2A) and depth (2B) scales and give sample responses from the English vocabulary data. Ambiguous responses were discussed among the raters, and when necessary, a third researcher resolved the disagreement.

## Analyses

The results are analyzed for PM 1–3. PM1 informs us of the relative performance of the children in each language at baseline,

which makes it possible to identify dominance and the impact of the intervention. PM2 immediately followed the intervention in HL (English) enabling identification of gains in both languages and CLI in Hebrew following the English intervention. PM3 immediately followed the intervention in SL (Hebrew), and enables identification of gains in both languages and CLI in English following the Hebrew intervention. PM4 is reported only in the descriptive statistics of the full corpus (not in further analyses) since it tested for maintenance and does not contribute to the discussion of immediate gains and CLI.

Descriptive statistics are reported for overall distribution of responses (numbers and percentages out of total responses) reporting full responses, partial responses, codeswitching and gestures, followed by  $\chi^2$  tests. Means and standard deviations calculated for the 0–1 scale were used in multivariate GLMs and one-way ANOVAs (with *post hoc* tests) to explore lexical breadth. Spearman correlational analyses and Linear Regressions were used to assess the impact of learner characteristics on gains to test for CLI. Crosstabs were used to compare the performance on the 0–1–2 scale to explore lexical depth. Manual analysis of the responses was used for the qualitative analysis to identify CLI in lexical depth.

To ensure procedural fidelity during the intervention sessions, a procedural reliability form consisting of all the steps in the intervention session was filled out. These records show that the procedure was implemented with over 99% reliability. To establish inter-rater reliability for the vocabulary task coding, 25% of the children's responses in each language were double coded by the authors and two graduate students who conducted the intervention and coded the data. The agreement percentage was calculated by dividing the number of discrepant scores over the total number of answers. Reliability for the coding was initially 94% for English and 92% for Hebrew, but after discussion and adjustments, full agreement was reached among raters.

## RESULTS

### Descriptive Statistics

The vocabulary knowledge task resulted in 4200 responses by the experimental group and 3528 responses by the control group. **Table 3** presents the overall distribution of the responses (numbers and percentages out of total responses) divided by Group and Language of Testing. The number of PMs is reported for each language in each group ( $N$  = number of children multiplied by four excluding missing children). Due to Covid19, one child in the experimental group was absent for one Hebrew PM, one child in the control group missed one English PM and four children in the control group missed one PM each. Total responses are the number of responses provided by the children and include full responses (depth score of 2), partial response (depth score of 1), codeswitches and gestures as well as wrong definitions and “I don't know”. The latter two categories are not presented in the table. Chi-squared ( $\chi^2$ ) tests were applied to the different types presented in the table for within group and between group comparisons.

**TABLE 2 |** Scoring for lexical breadth on the vocabulary knowledge task.

#### A.Vocabulary Breadth Scale

| Scores | Types of answers                                                                                                                                                                                                                      | Examples of answers*                                                                                                                                                                              |
|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 0      | - Wrong answer<br>- Gesture<br>- I don't know/no response                                                                                                                                                                             | - <i>SCRUB</i> : fall                                                                                                                                                                             |
| 1      | - Correct/complete answer<br>- Correct answer using a form of the word to define the word (but indicating complete comprehension)<br>- Close/partially correct answer/incomplete explanation<br>- Association<br>- Correct codeswitch | - <i>COLLECT</i> : bring together<br>- <i>SCRUB</i> : to scrub the floor with a brush until it's clean<br>- <i>SCRUB</i> : to clean (incomplete)<br>- <i>COOPERATE</i> : to be good (association) |

#### B.Vocabulary Depth Scale

| Scores | Types of answers                                                                                                                                                    | Examples of answers*                                                                                                                                    |
|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| 0      | - Wrong answer<br>- Gesture<br>- Codeswitch<br>- I don't know/no response                                                                                           | <i>DAMP</i> : really deep                                                                                                                               |
| 1      | - Correct answer using a form of the word for definition (but indicating comprehension)<br>- Close/partially correct answer/incomplete explanation<br>- Association | - <i>SCRUB</i> : He scrubbed the stain with a sponge<br>- <i>SCRUB</i> : to clean (incomplete answer)<br>- <i>TREMBLING</i> : you're cold (association) |
| 2      | - Correct/complete answer                                                                                                                                           | - <i>SCRUB</i> : to clean hard<br>- <i>COOPERATE</i> : to work together<br>- <i>TREMBLING</i> : shaking                                                 |

Target words are in all caps and italics.

**TABLE 3 |** Descriptive statistics of the full corpus of responses.

| Group             | Experimental     |                 | Within $\chi^2$ | Control          |                 | Within $\chi^2$ | Between groups $\chi^2$ |           |
|-------------------|------------------|-----------------|-----------------|------------------|-----------------|-----------------|-------------------------|-----------|
|                   | English [N = 88] | Hebrew [N = 87] |                 | English [N = 75] | Hebrew [N = 72] |                 | English                 | Hebrew    |
| Total responses   | 2112             | 2088            |                 | 1800             | 1728            |                 |                         |           |
| Full responses    | 546 (26%)        | 424 (20%)       | 18.18***        | 275 (15%)        | 274 (16%)       | ns.             | 65.53***                | 209.56*** |
| Partial responses | 425 (20%)        | 185 (9%)        | 107.2***        | 462 (26%)        | 128 (7%)        | 76.07***        | 17.03***                | ns.       |
| Code-switching    | 1 (0%)           | 10 (0.5%)       | 7.486**         | 0 (0%)           | 28 (1.6%)       | 27.09***        | ns.                     | 12.49***  |
| Gestures          | 22 (1%)          | 62 (3%)         | 19.9***         | 10 (0.5%)        | 30 (1.7%)       | 10.96***        | ns.                     | 6.11*     |

ns. = non-significant, \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .

Within group comparisons show English dominance for both groups. Within the experimental group, more full responses and more partial responses were provided in the English testing than in the Hebrew testing, while more codeswitches and gestures were used in the Hebrew testing. Within the control group, percentages of full responses were the same in English and Hebrew, but there were significantly more partial responses in English than in Hebrew. Like in the experimental group, more codeswitching and gestures were used in Hebrew than in English. Between group analyses demonstrate the impact of the intervention on both languages. The between group analysis shows that the experimental group gave more full responses in both languages and fewer partial responses in English, with no significant difference on partial responses in Hebrew. The experimental group used codeswitching less but gestures more in Hebrew when compared to the control group, with no difference in English for these measures.

In order to further explore the effect of bilingual narrative intervention (BINARI) that includes explicit vocabulary instruction on the lexical breadth and depth of the HL and SL lexicons of English-Hebrew preschool children, we then present the impact of the intervention on breadth and depth, comparing the experimental and control groups (Research Question 1). Then, focusing on immediate gains following the explicit vocabulary intervention, we explore whether learner characteristics such as bilingual experience and language proficiency predict performance and CLI for lexical breadth in the experimental group (Research Question 2). Finally, we focus on how cross-language influence from the HL to the SL and from the SL to the HL is manifested in lexical depth within the experimental group (Research Question 3).

## Impact of Lexical Intervention on Lexical Breadth in the HL and SL Lexicons

Data is reported for PM 1–3. PM4 for is not reported, as no significant changes were observed, showing maintenance of learning over time, but this does not contribute to the primary focus of the paper on immediate gains and CLI. **Table 4** presents the ratio of items the children were able to verbally define (including full and partial definitions and codeswitching) out of the 24 items tested in each language at the progress monitoring points for the experimental and control groups. PM1 shows the relative performance of the children in each language at baseline, which makes it possible to identify the impact of the intervention.

PM2 immediately followed the intervention in the HL (English). PM3 immediately followed the intervention in the SL (Hebrew).

A multivariate GLM with Group (Experimental, Control), Language of Testing (HL/English, SL/Hebrew) and PM point (1–3) as independent variables and vocabulary knowledge as the dependent variable yielded significant main effects for Group,  $F(1,228) = 9.06$ ,  $p = 0.003$ ,  $\eta^2 = 0.04$ , Language of Testing,  $F(1,228) = 53.48$ ,  $p < 0.001$ ,  $\eta^2 = 0.19$ , and PM point,  $F(2,228) = 8.21$ ,  $p < 0.001$ ,  $\eta^2 = 0.07$ , with no interaction between these variables. The experimental group demonstrated better performance than the control group. Performance in English was better than in Hebrew, and the children progressed across PM points.

Since half of the target items were taught in Hebrew and the other half in English, and children were tested on all target items and their translation equivalents in both languages, a second analysis was conducted, comparing performance on target items taught in the language of testing and their translation equivalents (those not taught in the language of testing) at the three PM points. A multivariate GLM with Language of Intervention (HL/English, SL/Hebrew) as a repeated factor within participants, Group (Experimental, Control) as independent variables between participants, Language of Testing (HL/English, SL/Hebrew), and PM point (1–3) as independent variables between items and vocabulary knowledge as the dependent variable was conducted to test the effect of language of intervention on growth in vocabulary breadth. Results demonstrated a significant main effect for Language of Intervention,  $F(1,228) = 41.41$ ,  $p < 0.001$ ,  $\eta^2 = 0.15$ , and significant interactions between (a) Language of Intervention and Language of Testing  $F(1,228) = 21.80$ ,  $p < 0.001$ ,  $\eta^2 = 0.08$ ; (b) Language of Intervention and PM point,  $F(2,228) = 3.09$ ,  $p = 0.047$ ,  $\eta^2 = 0.03$ ; (c) Language of Intervention, Language of Testing, and PM point,  $F(2,228) = 4.64$ ,  $p = 0.01$ .

**TABLE 4 |** Mean scores for vocabulary breadth in each progress monitoring session in each language (reported as ratios).

| Group        | N  | Language     | PM1         | PM2         | PM3         |
|--------------|----|--------------|-------------|-------------|-------------|
| Experimental | 22 | HL (English) | 0.33 (0.20) | 0.49 (0.22) | 0.53 (0.19) |
|              |    | SL (Hebrew)  | 0.24 (0.13) | 0.26 (0.17) | 0.33 (0.21) |
| Control      | 19 | HL (English) | 0.32 (0.14) | 0.37 (0.14) | 0.42 (0.15) |
|              |    | SL (Hebrew)  | 0.20 (0.13) | 0.25 (0.10) | 0.24 (0.11) |

Language, language of testing; HL, home language; SL, school language; PM, progress monitoring.

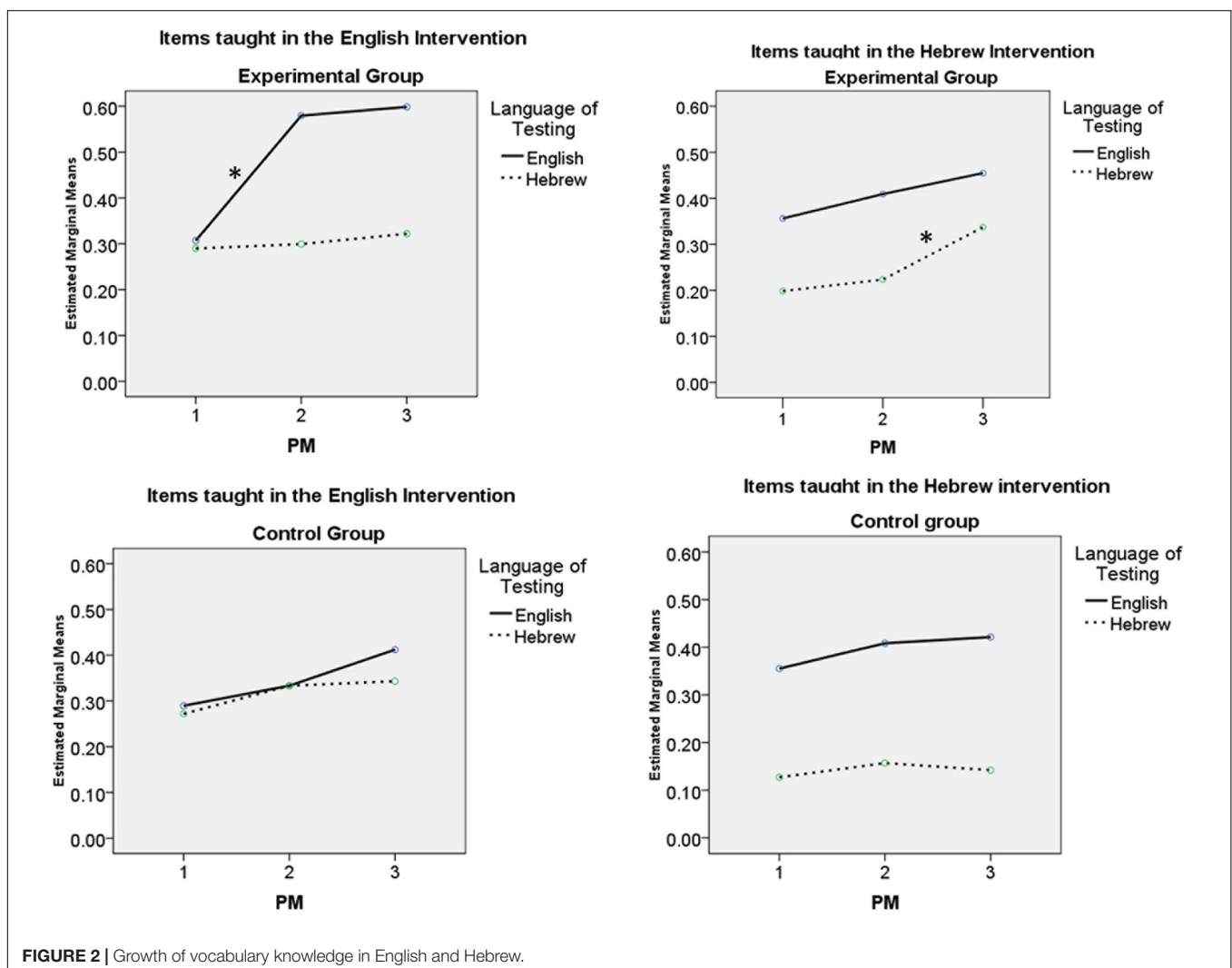
$\eta^2 = 0.04$ ; (d) Language of Intervention, Language of Testing, and Group (experimental/control)  $F(1,228) = 41.08$ ,  $p < 0.001$ ,  $\eta^2 = 0.15$ ; and (e) Language of Intervention, Language of Testing, Group, and PM point,  $F(2,228) = 5.59$ ,  $p = 0.004$ ,  $\eta^2 = 0.05$ .

**Figure 2** provides a visual representation of the growth of vocabulary knowledge in each language (presented as ratios of correct responses) from one PM point to the next as a function of language of intervention. **Figures 2A,B** present growth in the ratio of vocabulary knowledge in English and Hebrew testing for items taught in English (2a) and for items taught in Hebrew (2b) for the experimental group. **Figures 2C,D** presents the same information for the control group.

**Figure 2** and the GLM demonstrate that following intervention in HL/English, only children in the experimental group (**Figure 2A** vs. **Figure 2C**, English line from PM point 1 to 2) showed growth in vocabulary in HL/English for the target items presented in the intervention. Following intervention in SL/Hebrew (**Figure 2B** vs. **Figure 2D**, Hebrew line versus English line from PM point 2 to 3), children showed growth in SL/Hebrew vocabulary for those target

items taught in the intervention, with no change in the translation equivalents that were not subject to intervention. Children in the control group showed no such effects (**Figures 2C,D**, English line from PM point 1 to 2, Hebrew line from PM point 2 to 3). That is, bilingual narrative intervention with vocabulary instruction had a direct impact on lexical breadth in both the HL and SL lexicons of English-Hebrew preschool children, for the target items for which intervention was provided.

Further insights into the impact of the intervention within each language emerge from the immediate gains of the experimental group in each language following each intervention block. As can be expected following the results above, following the intervention in English significantly larger gains are observed in English ( $M = 0.27$ ) than in Hebrew ( $M = 0.05$ ) for target words that were taught in English,  $t(21) = 5.93$ ,  $p < 0.001$ , and following the intervention in Hebrew larger gains are observed in Hebrew ( $M = 0.11$ ) than in English ( $M = 0.02$ ) for target words that were taught in Hebrew, and the difference was nearly significant,  $t(21) = 2.04$ ,  $p = 0.054$ .



## Impact of Intervention on Lexical Depth in the HL and SL Lexicons

To understand the impact of the intervention on lexical depth, the quality of the definitions was explored for each item separately, using a 0–1–2 scoring scheme. Descriptively, improvement for the experimental group was observed in English for all target items taught in English after the English intervention (at PM2), and in Hebrew for all target items taught in Hebrew after the Hebrew intervention (at PM3). A Kruskal Wallis Test showed significant improvement in HL/English in the definitions, as a measure of lexical depth, between the first PM point at baseline and the second PM point after the English intervention for seven of the twelve items presented in the English intervention: damp ( $p = 0.017$ ), cooperate ( $p = 0.008$ ), narrow ( $p = 0.001$ ), wise ( $p = 0.008$ ), hidden ( $p = 0.007$ ), repair ( $p < 0.001$ ), and tremble ( $p < 0.001$ ). This was reflected by an increase in the number of complete answers. Such a significant improvement was observed in Hebrew following the Hebrew intervention for only one item (suffocate,  $p = 0.011$ ). No differences were found for the control group for any item in any language.

In sum, intervention in English affected the breadth and depth of lexical knowledge in English, while intervention in Hebrew mostly influenced breadth in Hebrew with very limited effect on depth of lexical knowledge. No differences between the PMs were observed for the control group on any of the measures. Thus, all further analyses to test for learner characteristic moderators and CLI (RQ2) focus only on the experimental group.

## Learner Characteristics and Cross-Language Influences

Our quantitative results so far show an impact of bilingual intervention on lexical breadth and depth only for the language of intervention, with no evidence for CLI. Yet, the large variability in learner characteristics within the group might have impacted the ability of the children to draw maximal benefits from

the intervention, transferring gains across languages. Learner characteristics include the bilingual experience (chronological age, age of onset of bilingualism and length of exposure) and proficiency in each language at the onset of the intervention (standardized tests, familiarity with the vocabulary introduced in the intervention at baseline).

To address the large variance in these variables and identify significant relationships between learner characteristics and gains, non-parametric correlational analyses (Spearman) were conducted for the gains between PM1 and PM2 and between PM2 and PM3 separately for each language of testing for target words that were taught in the particular language and their translation equivalents. Of special interest for CLI are the relations between learner characteristics and possible gains for translation equivalents (words that were taught in the other language), that is, gains in Hebrew between PM1 and PM2 for words taught in English, and gains in English between PM2 and PM3 for words taught in Hebrew. Due to the high correlation between length of exposure to the SL and AoB, only AoB was used in the correlations and the following regressions.

**Table 5** presents the correlation matrix of individual differences in chronological age, AoB, Hebrew proficiency, English proficiency and English vocabulary dominance at PM1 against HL and SL gains, separating target words from translation equivalents within each language. The upper right triangle of the matrix presents correlations for English while testing in English and the lower left triangle (italicized) presents correlations for Hebrew while testing in Hebrew. Numbers at the top (1–9) correspond to the measures in the first column on the left, but in the language of testing (English). Target words are words taught in the language of testing and translation equivalents are words taught in the other language. Significant correlations are in bold print.

Of the learner characteristics, age of onset of bilingualism, English proficiency at the onset of the intervention, and English vocabulary dominance (higher scores in English than Hebrew at baseline), presented significant correlations with gains in English

**TABLE 5 |** Correlation matrix of individual difference with HL and SL gains for target items and translation equivalents for each language separately.

| Hebrew testing                    | English testing |               |               |                       |               |               |               |                |        |
|-----------------------------------|-----------------|---------------|---------------|-----------------------|---------------|---------------|---------------|----------------|--------|
|                                   | 1               | 2             | 3             | 4                     | 5             | 6             | 7             | 8              | 9      |
| 1. Age in months                  | –               | 0.396         | 0.016         | 0.051                 | 0.356         | –0.129        | –0.152        | –0.146         | –0.060 |
| 2. AoB                            | <i>0.396</i>    | –             | –0.353        | 0.017                 | –0.065        | <b>0.423*</b> | –0.300        | 0.199          | –0.169 |
| 3. Hebrew proficiency             | <i>0.016</i>    | <i>–0.353</i> | –             | –0.114                | –0.369        | 0.103         | 0.343         | 0.168          | 0.186  |
| 4. English proficiency            | <i>0.051</i>    | <i>0.017</i>  | <i>–0.114</i> | –                     | 0.127         | <b>0.434*</b> | –0.141        | 0.177          | –0.109 |
| 5. English dominance at PM1       | <i>0.356</i>    | <i>–0.065</i> | <i>–0.369</i> | <i>0.127</i>          | –             | –0.136        | –0.398        | <b>–0.471*</b> | 0.009  |
| 6. Target PM1_PM2                 | <i>0.280</i>    | <i>–0.157</i> | <i>0.231</i>  | <b><i>–0.423*</i></b> | <i>0.168</i>  | –             | –0.244        | 0.330          | –0.010 |
| 7. Target PM2_PM3                 | <i>0.318</i>    | <i>0.127</i>  | <i>0.210</i>  | <b><i>0.613**</i></b> | <i>0.174</i>  | <i>–0.238</i> | –             | 0.281          | 0.112  |
| 8. Translation equivalent PM1_PM2 | <i>0.178</i>    | <i>0.138</i>  | <i>0.344</i>  | <i>0.170</i>          | <i>0.120</i>  | <i>0.149</i>  | <b>0.445*</b> | –              | –0.083 |
| 9. Translation equivalent PM2_PM3 | <i>–0.027</i>   | <i>–0.016</i> | <i>0.039</i>  | <i>–0.150</i>         | <i>–0.120</i> | <i>–0.006</i> | <i>0.158</i>  | <i>–0.118</i>  | –      |

\* $p < 0.05$  and \*\* $p = 0.002$ .

The upper right triangle in the matrix presents correlations for English while testing in English and the lower left triangle (italicized) presents correlations for Hebrew while testing in Hebrew. Significant correlations are in bold print. Numbers at the top (1–9) correspond to the measures in the first column on the left, but in the language of testing (English). AoB, age of onset of bilingualism; Hebrew proficiency, raw Goralnik score; English proficiency, CELF CLS; English dominance, relative familiarity with the vocabulary at baseline; PM, progress monitoring; Target words, words taught in the language of testing; Translation equivalents, words taught in the other language.

and Hebrew. Of particular interest for CLI is the correlation between gains in Hebrew for translation equivalents of English words at PM2 and gains in Hebrew on target words in PM3 (Spearman  $Rho = 0.445$ ,  $p = 0.038$ ). In an attempt to tease apart the relative contribution of each of these three predictors – AoB, English proficiency, and English vocabulary dominance – to the gains in English and Hebrew, linear regression analyses were conducted on the gains observed in the two languages with all predictors entered simultaneously. Three models came out significant.

The first model explores which learner characteristics predict direct gains from the English intervention. In the first regression, these three variables accounted for 36% of the variance in gains in English between PM1 and PM2 on target items that were taught in English,  $F(3,18) = 3.38$ ,  $p = 0.04$  (Model 1 in Table 6). Children with later AoB and those who started out with higher English proficiency gained more in English from the English intervention.

The second model aims to find out if gains are extended beyond the taught items to the translation equivalents. In the second regression, the same three variables accounted for 45% of the variance in gains in English between PM1 and PM2 on items that were not taught yet (i.e., English translation equivalent of Hebrew target words taught after PM2 in the Hebrew intervention),  $F(3,18) = 4.96$ ,  $p = 0.01$  (see Model 2 in Table 6). Here, children who demonstrated Hebrew vocabulary dominance (i.e., they began with lower English vocabulary scores than Hebrew vocabulary scores at the onset of the intervention) gained more in English translation equivalents for items not taught yet. No predictors were found for gains in Hebrew between PM1 and PM2 that would count as predictors of CLI.

In the third model, the three variables accounted for 32% of the variance in gains in Hebrew between PM2 and PM3 on

target items that were presented in the Hebrew intervention, but these gains were not significant  $F(3,18) = 2.82$ ,  $p = 0.07$  (see Model 3 in Table 6). Here children with better English proficiency gained from the intervention in Hebrew. No predictors were found for gains in Hebrew translation equivalents or in English between PM2 and PM3 that would count as predictors of CLI. That is, only gains in HL/English between PM1 and PM2 yielded significant models.

## Cross-Language Influence on Lexical Depth

Since the quantitative analyses did not show evidence of CLI, a qualitative analysis of the children's responses was used to explore cross-language influence from the HL to the SL and from the SL to the HL for lexical depth. Two different types of indications of CLI were noted, one focused on the lexical item and one on the definition taught in the intervention. The first type was transfer of lexical information about a word in one language to a word in the other language. The second type was transfer of definitions taught in one language to the definitions in the other language. Both phenomena were documented in both directions, i.e., from HL to SL and from SL to HL.

An example of a target word that elicited both types of CLI (semantic and definitional) is the word "float." In English, "float" includes the meaning of floating in the air or floating on the surface of water, but the Hebrew word "*tsaf*" can only mean floating on water. (To talk about floating in the air, one would use the word "*af*" = fly). The children were taught the word only in Hebrew, and the definition provided was "to stay on top of the water" (in Hebrew). When asked to define the Hebrew word "float," six of the children initially overextended the definition of float from English and responded "fly" (which was scored as

**TABLE 6 |** AoB, English proficiency, and English vocabulary dominance as predictors in regression models.

| Predictor                                                                                                                        | <i>B</i> | Standard error | $\beta$ | <i>t</i> | <i>p</i> |
|----------------------------------------------------------------------------------------------------------------------------------|----------|----------------|---------|----------|----------|
| <b>Model 1: Predictors of gains in English for English target words at PM2</b>                                                   |          |                |         |          |          |
| AoB                                                                                                                              | 0.005    | 0.002          | 0.423   | 2.244    | 0.038    |
| English proficiency                                                                                                              | 0.007    | 0.003          | 0.440   | 2.280    | 0.035    |
| English vocabulary dominance                                                                                                     | −0.054   | 0.164          | −0.063  | −0.329   | 0.746    |
| $R^2 = 0.36F(3,21) = 3.39$ , $p = 0.04$                                                                                          |          |                |         |          |          |
| <b>Model 2: Predictors of gains in English for English translation equivalent (of Hebrew target words not taught yet) at PM2</b> |          |                |         |          |          |
| AoB                                                                                                                              | 0.002    | 0.002          | 0.198   | 1.135    | 0.271    |
| English proficiency                                                                                                              | 0.006    | 0.003          | 0.366   | 2.048    | 0.055    |
| English vocabulary dominance                                                                                                     | −0.534   | 0.158          | −0.606  | −3.392   | 0.003    |
| $R^2 = 0.45F(3,21) = 4.95$ , $p = 0.01$                                                                                          |          |                |         |          |          |
| <b>Model 3: Predictors of gains in Hebrew for Hebrew target words at PM3</b>                                                     |          |                |         |          |          |
| AoB                                                                                                                              | 0.001    | 0.002          | 0.109   | 0.562    | 0.581    |
| English proficiency                                                                                                              | 0.009    | 0.003          | 0.549   | 2.762    | 0.013    |
| English vocabulary dominance                                                                                                     | 0.027    | 0.176          | 0.031   | 0.154    | 0.879    |
| $R^2 = 0.32F(3,21) = 2.82$ , $p = 0.07$                                                                                          |          |                |         |          |          |

AoB, age onset of bilingualism.

correct in English but incorrect in Hebrew). However, following the Hebrew intervention, none of the children responded “af” (fly) as an answer in Hebrew (though four still did so in English). References to water increased (in both languages, though more so in Hebrew than in English), and six of the children gave more precise definitions as a result of intervention in Hebrew (e.g., “to stay on top of the water,” instead of “going in the water”). This word shows that the children transferred concepts learned in one language to the other language, as well as definitions learned, even without any explicit instruction regarding the crosslinguistic connections.

There were other examples of the children possibly transferring concepts from one language to the other. The word “search,” taught in English, elicited Hebrew definitions of “look” or “see” for three children after the English intervention, which were close to the English definition of “look for.” Of note, “to look” in English was considered a partially correct definition, but in Hebrew “*lhistakef*” (“to look”) was scored as incorrect/less appropriate. One child, who when asked for a Hebrew definition of “*xazak*” (“powerful”), a word which had been taught in English, code-switched into English and said “powerful.” Generally, one would expect children of this age to use the more frequent, lexically unmarked (Raichlin et al., 2018) word “strong” in their codeswitching, so the choice of “powerful” seems to indicate a connection between the semantic representations of the words. Another example is “hidden,” taught in the English intervention. In Hebrew, the word “*nistar*” (hidden) has no phonological or morphological relation to “*l’haxbi*” (hide) as hidden and hide do in English. However, after being taught the word “hidden” in the English intervention, three children gave definitions during Hebrew testing that included the notion of hiding, which might indicate a connection between the English and Hebrew representations.

An indication of definitional CLI is exemplified in the word “cooperate” in English, defined in the intervention as “to work together.” Many of the children knew something about this word in Hebrew before the English intervention; they gave partially correct responses and associations in Hebrew such as “*shatef*” (to share), “*marshet*...” (to allow other kids to play with them), or “*leshatef et haxaverim*” (to include their friends). (This may be due to the fact that “share” and “cooperate” in Hebrew have a common root). However, following the intervention, 10 of the children improved their Hebrew definitions by giving responses in Hebrew that were similar to the English definition they were taught. Thus, the children improved their definitions in a given language by using the definitions taught in the other language, implying that CLI occurred.

## DISCUSSION

The current study explored the effects of bilingual narrative intervention with embedded vocabulary instruction on lexical breadth and depth in English-Hebrew bilingual kindergarteners. Results showed that the intervention increased the breadth and depth of the experimental children’s vocabulary in their HL/English and lexical breadth in the SL/Hebrew. In terms of modulating factors, no effect was found for the modulating

factors on CLI. Better HL performance was correlated with later AoB (later acquisition of SL) and higher HL language proficiency scores. The children with higher HL proficiency also responded better to the SL/Hebrew intervention, gaining more in lexical breadth than those with lower English proficiency. Children dominant in SL/Hebrew vocabulary at the outset of the study also gained more from the HL/English intervention. Finally, bidirectional CLI was found for semantic information in a qualitative analysis of children’s responses. Since no gains were detected in the control group, the following discussion focuses only on the experimental group, relating first to the findings for lexical breadth and depth, then to CLI, and finally to the influence of bilingual learner characteristics.

## Lexical Breadth and Depth

Gains following bilingual vocabulary intervention in this study are in line with previous research on vocabulary interventions, which show that they are effective for preschool children (Marulis and Neuman, 2010) and that bilingual interventions lead to progress in both languages (Restrepo et al., 2013). In the present intervention, the children gained relatively more in HL/English than in SL/Hebrew. For most of these children, SL/Hebrew was the weaker language as seen in their Hebrew proficiency scores, where nine of the children scored more than 1.25 SD below the mean. Future interventions should examine the effect of beginning the intervention in the school language, Hebrew, to compare the effects of beginning with the home vs. school language.

The findings for depth also differed for the two languages. For HL/English, the intervention led to an increase in the depth of the children’s knowledge while for SL/Hebrew the effect on depth was more limited. This can be attributed to the different proficiency levels of the children’s two languages. As explained by the Revised Hierarchical Model (Kroll and Stewart, 1994; Kroll and Ma, 2018), improved proficiency leads to more direct access to shared conceptual representations of words, with these stronger links resulting in enhanced performance on language tasks. This could explain the differential performance in the two languages of the children, with better performance in HL/English, the dominant language of this group.

## Crosslinguistic Influence

Analysis of semantic CLI across languages showed evidence for transfer of knowledge from the intervention, resulting in more precise definitions. We interpret this as a reflection of deeper lexical knowledge. This phenomenon was documented bidirectionally, both from HL to SL and from SL to HL. Semantic CLI finds support in the Revised Hierarchical Model (Kroll and Stewart, 1994; Kroll and Ma, 2018) which posits shared semantic representations between words and concepts and lexical connections between words, and activation of the lexicons of both languages. MacWhinney’s Unified Competition Model also explains this bidirectional CLI, as similar features in languages may influence each other. The findings of the present study provide details about how these connections and representations change as a result of bilingual narrative intervention with a focus on vocabulary. As reported above, the word “float,” which includes a broader semantic range in English (“float” can be used

with air and water) than in Hebrew (“*tsaf*” = float can be used with water only) resulted in overgeneralizations in Hebrew due to CLI. Initially, children may assume semantic similarity or even one-to-one semantic mapping between translation equivalents until they are taught otherwise (Jarvis and Pavlenko, 2008). Indeed, following the intervention, where the children are given precise information, they improved and crosslinguistically transferred the lexical knowledge taught in the other language. The present study adds to others which show evidence for these connections (Singh, 2014; Floccia et al., 2020).

The results of the present study also provide support for the activation of both languages in bilinguals (e.g., Singh, 2014; Degani et al., 2018; Floccia et al., 2020). Once the child has heard a word in one language, both lexicons can be activated and the child can draw from the conceptual network, which contains words from both languages. As Mancilla-Martinez and Vagh (2013) note, information about concepts tied to vocabulary known in one language may facilitate acquisition of translation equivalents and transfer of linguistic information. Previous studies have reported crosslinguistic activation to be stronger from the dominant language to the non-dominant language (Singh, 2014), and bidirectional connections have been found as well (Floccia et al., 2020). In the current study, crosslinguistic associations were seen in both directions: from HL to SL and from SL to HL. Where CLI was found as a result of intervention, the shared mechanism underlying all language learning may have facilitated the CLI of lexical information (Cummins, 1981, 2008).

While gains for breadth did not show CLI from the language of intervention to the other language, children did produce codeswitching in their responses, which in the present context can be considered a form of CLI. Codeswitching was observed more in Hebrew than in English and more among the control group than the experimental group (see Table 3). These findings show that CLI was modulated by the intervention, which increased the children’s use of full or partial responses, making codeswitching redundant. CLI was also modulated by language dominance.

In spite of the evidence of transfer of lexical knowledge crosslinguistically, which deepened the children’s knowledge of words that had been partially acquired, we did not find evidence of increased lexical breadth in a given language as a result of intervention in the other language for the experimental group as a whole. However, the children who did show gains in Hebrew for translation equivalents following the English intervention (at PM2), showed further gains in Hebrew target words following the Hebrew intervention (at PM3). This suggests that the children who rely on CLI at PM2 benefit more from the intervention. That is, in the quantitative analysis of CLI, words taught in one language did not automatically lead to significant gains on those words in the other language for all children. It seems that the improvements in the children’s responses as a result of CLI occurred more in cases where the children knew something about a word, but learned more as a result of the intervention (improvement in depth), thus limiting the number of potential words influenced by instruction in the other language. Indeed, in some cases, it was possible for a child’s score to remain stable while still being affected by CLI, for example in cases where

responses from earlier PMs received maximal scores for lexical breadth (which included partial or complete knowledge), but then improved even further after the intervention, although it was not reflected in an increase in the score.

Although some evidence has been found in previous studies for CLI involving lexical breadth (Armon-Lotem et al., 2020), many scholars have noted that lexical items are presumed to be learned one at a time in each language (David and Wei, 2008; Bialystok et al., 2010), and thus are unlikely to transfer automatically across languages without explicit connections being made, unless they are cognates (August et al., 2005). In the current study, if explicit connections had been made across languages [e.g., by providing translations of words learned, or by pointing out associations across languages, as suggested by Lugo-Neris et al. (2015)] it is possible that the children would have learned more words in the other language. It is also possible that additional follow-up questions for each word may have elicited more information about the children’s knowledge and would have revealed more evidence of growth as a result of CLI. A more differentiated scale may also have captured more growth in the children’s knowledge. Moreover, the scale was designed to measure semantic content, not definitional form. Since previous research has shown that formal definition is one aspect of lexical knowledge that transfers (Ordóñez et al., 2002; Pham et al., 2018), it is feasible that the design of the scale and the scoring did not allow all the transfer effects to be detected. Future research should address these issues.

## Bilingual Experience

Our quantitative results showed impact of bilingual intervention on lexical breadth and depth only for the language of intervention. A possible reason for this finding could have been variability within the group in their bilingual experience (chronological age, age of onset of bilingualism, and length of exposure) and variability in proficiency in each language at the onset of the intervention (standardized tests, familiarity with the 24 vocabulary items introduced in the intervention). These factors might have impacted the ability of the child to draw maximal benefits from the intervention.

The regression analyses showed that children with later AoB and/or higher English proficiency benefited more from the intervention in HL/English and showed greater improvement in lexical breadth (Model 1), which may be due to their larger HL lexicon. Moreover, following intervention in Hebrew (PM2 to PM3), these same children showed higher gains than those with lower English proficiency in English for English translation equivalents (model 2) and in Hebrew for the Hebrew target words taught in the intervention (Model 3). That is, the stronger one’s HL, the larger one’s direct (Models 1 and 3) and indirect (Model 2) gains from the intervention. This raises the question as to why. One answer may be related to the process by which words are acquired. At PM1 and PM2, the children are exposed to unfamiliar concepts they do not recognize but acquisition has begun. Once they build foundations, even in another language, their strong HL foundation allows them to make larger SL gains once intervention in the SL is internalized (Cummins, 1979, 1981; Lugo-Neris et al., 2015). Thus, the fact that the vocabulary

intervention was provided first in the stronger language may have enhanced their word learning skills and provided a basis for learning in the weaker language.

Moreover, those children whose English vocabulary was smaller than their Hebrew vocabulary showed an increase in their English vocabulary breadth for the translation equivalents that were not taught yet (Model 2). Similarly, children with low proficiency in English demonstrated an increase for these same words in Hebrew before they were taught, showing that they might have improved their metalinguistic skills and better understood how to define words. That is, these children generalized the abilities from the items they learnt to those they did not learn in both languages. The gain in English can further be attributed to the fact that they had more room to grow since they started at a lower point. This finding is in contrast to the Matthew effect, where the rich get richer (Stanovich, 1986); those who are relatively “poorer” in the HL vocabulary gain more in the HL, suggesting that intervention focused on teaching vocabulary in a context-rich environment with multiple exposures to words and opportunities for practice may be a way of counteracting the Matthew effect and boosting the weaker language. However, since all of the children had the same order of intervention (HL first, SL second), the mechanism here is to be further researched. As mentioned above, future research manipulating the order of language of intervention will help to clarify this issue.

## Limitations

The population in this study consisted of typically developing children from mid-high SES levels with normal language proficiency and IQ scores and no comorbidities. It is therefore necessary in the future to examine whether these results are applicable to other demographics with various risk factors. Nevertheless, given the success of the intervention with this population, it is possible that this intervention program may have promise for helping atypical bilingual populations, such as children with developmental language disorders (DLD), to strengthen their language skills. Given the known difficulties of this population in expressive language, it is possible that a more structured task such as asking questions may be more fruitful for examining depth of knowledge. In addition, this study was focused on exploring CLI. Different words were intentionally taught in the two languages to test for CLI of conceptual knowledge. This specific design feature does not allow for a direct comparison between the items of the two interventions in the two languages, which is a potential limitation of the current study.

Another limitation relates to the data collection. Given that the onset of the Covid-19 pandemic coincided with a portion of the data collection, some of the data were collected online. Based on observation of the children in addition to their self-reports about their online experience as compared to the face-to-face sessions, it was concluded that the online mode did not alter children’s performance. In addition, the experience was generally positive and raises the possibility of continuing to use this method in the future. This result is in line with other studies that found children’s online and offline performance on language assessments to be comparable (Waite et al., 2010; Ciccia et al., 2011; Sutherland et al., 2017; Manning et al., 2020), but more research is needed to confirm these findings.

## Conclusion

The present study examines the effects of bilingual narrative-vocabulary intervention on lexical breadth and depth and evidence for CLI. The study shows that bilingual narrative intervention with vocabulary instruction may be efficacious for improving the lexical breadth and depth of bilingual kindergarten children, which may be critical for their future academic success (Dickinson et al., 2003; Han, 2012; Kieffer, 2012). Although no quantitative evidence for CLI was observed, this study provides additional evidence for the simultaneous activation of lexicons in both languages when using either language (Kroll and Stewart, 1994; Singh, 2014; De Anda and Friend, 2020), leading to CLI of lexical depth. This has implications for planning intervention, as it suggests that both languages may be used to facilitate lexical growth. In addition, as in previous research (Rolstad et al., 2005; Restrepo et al., 2013; Mendez et al., 2015), support of the HL did not hinder the development of the school language, and was efficacious in stimulating growth in school language knowledge, due to CLI. Later AoB (later acquisition of SL) and higher HL language proficiency were associated with better HL and SL performance, suggesting that a strong basis in the HL may enhance linguistic outcomes, providing further support for maintaining and strengthening the HL of bilingual children. Hence, this type of intervention may be used to support both languages of dual language learners, especially those whose HL is not taught in school, and who may be at risk for HL attrition (Ardila et al., 2019). The results here contribute to the body of research about the development of the lexical-semantic networks of bilingual children and highlight the importance of strengthening both the HL and SLs, as well as a method for doing so.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Bar-Ilan University Institutional Review Board on Humanities and the Israeli Ministry of Education Chief Scientist. Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

ML, CA, JW, and SA-L were responsible for the conception, analysis, and interpretation of the work, and drafted and revised the article include intellectual content. CA was accountable for the integrity and accuracy of the work. SA-L and JW was approved the article for publication of content. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

This research has been supported by Israel Science Foundation Grant ISF 5388/19 to CA and JW. We would like to thank Prof. Adelaida Restrepo from Arizona State University for her support in developing the research design during her visit to the Language Development Lab at Bar-Ilan University using a Fulbright Senior Scholars Grant. We would also like to thank the research assistants at the Language Development Lab for data

collection, as well as the children and parents who agreed to participate in this research.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.671928/full#supplementary-material>

## REFERENCES

- Altman, C., Armon-Lotem, S., Fichman, S., and Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English-Hebrew bilingual preschool children with and without specific language impairment. *Appl. Psycholinguist.* 37, 165–193 doi: 10.1017/s0142716415000466
- Altman, C., Goldstein, T., and Armon-Lotem, S. (2018). Vocabulary, metalinguistic awareness and language dominance among bilingual preschool children. *Front. Psychol.* 9:1953
- Ardila, A., Rosselli, M., Ortega, A., Lang, M., and Torres, V. L. (2019). Oral and written language abilities in young Spanish/English bilinguals. *Int. J. Biling.* 23, 296–312 doi: 10.1177/1367006917720089
- Armon-Lotem, S. (2014). Between L2 and SLI: inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. *J. Child Lang.* 41, 3–33 doi: 10.1017/s0305000912000487
- Armon-Lotem, S., and Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *Int. J. Lang. Commun. Disord.* 51, 715–731. doi: 10.1111/1460-6984.12242
- Armon-Lotem, S., Restrepo, M. A., Lipner, M., Ahituv-Shlomo, P., and Altman, C. (2020). Vocabulary gains in bilingual narrative intervention. *Lang. Speech Hear. Serv. Sch.* 52, 436–448 doi: 10.1044/2020\_lshss-19-00117
- Armon-Lotem, S., Rose, K., and Altman, C. (2021). The development of English as a heritage language: the role of chronological age and age of onset of bilingualism. *First Lang.* 41, 67–89 doi: 10.1177/0142723720929810
- August, D., Carlo, M., Dressler, C., and Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learn. Disabil. Res. Pract.* 20, 50–57 doi: 10.1111/j.1540-5826.2005.00120.x
- Benelli, B., Belacchi, C., Gini, G., and Lucangeli, D. (2006). “To define means to say what you know about things”: the development of definitional skills as metalinguistic acquisition. *J. Child Lang.* 33, 71–97 doi: 10.1017/s0305000905007312
- Bernardini, P., and Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: the ivy hypothesis. *Biling. Lang. Cogn.* 7, 49–69 doi: 10.1017/s1366728904001270
- Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Can. J. Exp. Psychol.* 60, 68–79 doi: 10.1037/cjep2006008
- Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Biling. Lang. Cogn.* 13, 525–531 doi: 10.1017/s1366728909990423
- Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., and Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition* 140, 122–134 doi: 10.1016/j.cognition.2015.03.013
- Blackwell, A. A. (2005). Acquiring the English adjective lexicon: relationships with input properties and adjectival semantic typology. *J. Child Lang.* 32, 535–562 doi: 10.1017/s0305000905006938
- Carey, S. (1978). “The child as word learner,” in *Linguistic Theory and Psychological Reality*, eds M. Halle, J. Bresnan, & A. Miller (Cambridge, MA: MIT Press), 264–293.
- Carlo, M. S., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D., et al. (2004). Closing the gap: addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *J. Educ.* 189, 57–76 doi: 10.1177/0022057409189001-205
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguist. Approaches Biling.* 1, 318–345 doi: 10.1075/lab.1.3.05cho
- Ciccia, A. H., Whitford, B., Krumm, M., & McNeal, K. (2011). Improving the access of young urban children to speech, language and hearing screening via telehealth. *J. Telemed. Telecare* 17, 240–244 doi: 10.1258/jtt.2011.100810
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Rev. Educ. Res.* 49, 222–251 doi: 10.3102/00346543049002222
- Cummins, J. (1981). “The role of primary language development in promoting educational success for language minority students,” in *Schooling and Language Minority Students: a Theoretical Framework*, ed. California State Department of Education (Los Angeles, CA: National Dissemination and Assessment Center), 3–49.
- Cummins, J. (2008). “Teaching for transfer: challenging the two solitudes assumption in bilingual education,” in *Encyclopedia of Language and Education*, ed. N. H. Hornberger. Boston: Springer
- Dam, Q., Pham, G. T., Pruitt-Lord, S., Limon-Hernandez, J., and Goodwiler, C. (2020). Capitalizing on cross-language similarities in intervention with bilingual children. *J. Commun. Disord.* 87:106004 doi: 10.1016/j.jcomdis.2020.106004
- David, A., and Wei, L. (2008). Individual differences in the lexical development of French-English bilingual children. *Int. J. Biling. Educ. Biling.* 11, 598–618 doi: 10.2167/beb478.0
- De Anda, S., and Friend, M. (2020). Lexical-semantic development in bilingual toddlers at 18 and 24 Months. *Front. Psychol.* 11:3500
- DeAnda, S., Poulin-Dubois, D., Zesiger, P., and Friend, M. (2016). Lexical processing and organization in bilingual first language acquisition: guiding future research. *Psychol. Bull.* 142, 655–667 doi: 10.1037/bul0000042
- Degani, T., Prior, A., and Hajajra, W. (2018). Cross-language semantic influences in different script bilinguals. *Bilingualism* 21, 782–804. doi: 10.1017/s1366728917000311
- Dickinson, D. K., McCabe, A., & Anastasopoulos, L. (2003). “A framework for examining book reading in early childhood classrooms,” in *On Reading Books to Children: Parents and Teachers*, eds A. Van Kleeck, S. A. Stahl, & E.B. Bauer (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 95–113.
- Ebert, K. D., Kohnert, K., Pham, G., Disher, J. R., and Payesteh, B. (2014). Three treatments for bilingual children with primary language impairment: examining cross-linguistic and cross-domain effects. *J. Speech Lang. Hear. Res.* 57, 172–186 doi: 10.1044/1092-4388(2013)12-0388
- Floccia, C., Delle Luche, C., Lepadatu, I., Chow, J., Ratnage, P., and Plunkett, K. (2020). Translation equivalent and cross-language semantic priming in bilingual toddlers. *J. Mem. Lang.* 112:104086 doi: 10.1016/j.jml.2019.104086
- Foroodi-Nejad, F., and Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian-English bilinguals. *Biling. Lang. Cogn.* 12, 411–427 doi: 10.1017/s1366728909990241
- Goodrich, J. M., Lonigan, C. J., Kleuver, C. G., and Farver, J. M. (2016). Development and transfer of vocabulary knowledge in Spanish-speaking language minority preschool children. *J. Child Lang.* 43, 969–992 doi: 10.1017/s030500091500032x
- Goralnik, E. (1995). *Language Screening Test for Hebrew-Speaking Preschool Children*. Israel: Gai Agencies.
- Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., and Nesbitt, K. T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Read. Res. Q.* 51, 181–198. doi: 10.1002/trq.130

- Hai Weiss, A. (2020). "Lexical-semantic organization in monolingual and bilingual Hebrew speaking children: evidence from a word association task," in *An Anthology of Bilingual Child Phonology*, eds E. Babatouli & M. J. Ball. Bristol: Multilingual Matters.
- Håkansson, G., Eva-Kristina, S., and Ulrika, N. (2003). Measuring language proficiency in bilingual children: Swedish-Arabic bilingual children with and without language impairment. *Ling* 41, 255–288
- Han, W. J. (2012). Bilingualism and academic achievement. *Child Dev.* 83, 300–321 doi: 10.1111/j.1467-8624.2011.01686.x
- Hoff, E., and Core, C. (2013). Input and language development in bilingually developing children. *Semin. Speech Lang.* 34, 215–226 doi: 10.1055/s-0033-1353448
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27 doi: 10.1017/s0305000910000759
- Iluz-Cohen, P., and Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Biling Lang. Cogn.* 16, 884–899 doi: 10.1017/s1366728912000788
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York: Routledge.
- Johnson, C. J., and Anglin, J. M. (1995). Qualitative developments in the content and form of children's definitions. *J. Speech Lang. Hear. Res.* 38, 612–629 doi: 10.1044/jshr.3803.612
- Kan, P. F., and Kohnert, K. (2012). A growth curve analysis of novel word learning by sequential bilingual preschool children. *Biling. Lang. Cong.* 15, 452–469 doi: 10.1017/s1366728911000356
- Kieffer, M. J. (2012). Early oral language and later reading development in Spanish-speaking English language learners: evidence from a nine-year longitudinal study. *J. Appl. Dev. Psychol.* 33, 146–157 doi: 10.1016/j.appdev.2012.02.003
- Korat, O., Levin, I., Atishkin, S., and Turgeman, M. (2014). E-book as facilitator of vocabulary acquisition: support of adults, dynamic dictionary and static dictionary. *Read. Writ.* 27, 613–629 doi: 10.1007/s11145-013-9474-z
- Kroll, J. F. and Ma, F. (2018). "The bilingual lexicon," in *Blackwell Handbooks in Linguistics. the Handbook of Psycholinguistics*, eds E. M. Fernandez & H. S. Cairns (Oxford: Blackwell), 294–319.
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174 doi: 10.1006/jmla.1994.1008
- Kupisch, T. (2008). "Dominance, mixing and cross-linguistic influence: on their relation in bilingual development," in *First Language Acquisition of Morphology and Syntax: Perspectives Across Languages and Learners*, eds P. Guijarro-Fuentes, P. Larrañaga & J. Clibbens (Amsterdam: John Benjamins), 209–234. doi: 10.1075/lald.45.09kup
- Lugo-Neris, M. J., Bedore, L. M., and Peña, E. D. (2015). Dual language intervention for bilinguals at risk for language impairment. *Semin. Speech Lang.* 36, 133–142 doi: 10.1055/s-0035-1549108
- MacSwan, J., and Rolstad, K. (2005). Modularity and the facilitation effect: psychological mechanisms of transfer in bilingual students. *Hisp. J. Behav.* 27, 224–243 doi: 10.1177/0739986305257173
- MacWhinney, B. (2005). "A unified model of language acquisition," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. Kroll & A. De Groot, (Oxford: Oxford University Press), 49–67.
- Mancilla-Martinez, J., and Vagh, S. B. (2013). Growth in toddlers' Spanish, English, and conceptual vocabulary knowledge. *Early Child. Res. Q.* 28, 555–567 doi: 10.1016/j.ecresq.2013.03.004
- Manning, B. L., Harpole, A., Harriott, E. M., Postolowicz, K., and Norton, E. S. (2020). Taking language samples home: feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. *J. Speech Lang. Hear. Res.* 63, 3982–3990
- Marulis, L. M., and Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Rev. Educ. Res.* 80, 300–335 doi: 10.3102/0034654310377087
- Meir, N., Walters, J., and Armon-Lotem, S. (2017). Bidirectional morpho-syntactic transfer: the role of age of onset and language proficiency. *Linguist. Approaches Biling.* 7, 514–553
- Mendez, L. I., Crais, E. R., Castro, D. C., and Kainz, K. (2015). A culturally and linguistically responsive vocabulary approach for young Latino dual language learners. *J. Speech Lang. Hear. Res.* 58, 93–106 doi: 10.1044/2014\_jslhr-12-0221
- Ordóñez, C. L., Carlo, M. S., Snow, C. E., and McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: which vocabulary skills transfer? *J. Educ. Psych.* 94, 719–728 doi: 10.1037/0022-0663.94.4.719
- Ouellette, G. P. (2006). What's meaning got to do with it: the role of vocabulary in word reading and reading comprehension. *J. Educ. Psych.* 98, 554–566 doi: 10.1037/0022-0663.98.3.554
- Paradis, J. (2010). The interface between bilingual development and specific language impairment. *Appl. Psycholinguist.* 31, 227–252
- Paradis, J. (2011). Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguist. Approaches. Biling.* 1, 213–237 doi: 10.1075/lab.1.3.01par
- Pearson, B. Z., Fernández, S. C., and Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43, 93–120 doi: 10.1111/j.1467-1770.1993.tb00174.x
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Appl. Psycholinguist.* 18, 41–58 doi: 10.1017/s0142716400009863
- Pearson, B. Z., Umbel, V. C., Andrews de Flores, P., & Cobo-Lewis, A. (1999). Measuring cross-language vocabulary in childhood bilinguals at different stages of development. *Paper presented at Texas Research Symposium on Language Diversity*, Austin, TX.
- Petersen, D. B., and Spencer, T. D. (2016). Using narrative intervention to accelerate canonical story grammar and complex language growth in culturally diverse preschoolers. *Top. Lang. Disord.* 36, 6–19 doi: 10.1097/tld.0000000000000078
- Pham, G., Donovan, D., Dam, Q., and Contant, A. (2018). Learning words and definitions in two languages: what promotes cross-language transfer? *Lang. Learn.* 68, 206–233 doi: 10.1111/lang.12274
- Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., and Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *Int. J. Biling.* 17, 57–70 doi: 10.1177/1367006911431198
- Proctor, C. P., Silverman, R. D., Harring, J. R., and Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Read. Writ.* 25, 1635–1664 doi: 10.1007/s11145-011-9336-5
- Raichlin, R., Walters, J., and Altman, C. (2018). Some wheres and whys in bilingual codeswitching: directionality, motivation and locus of codeswitching in Russian-Hebrew bilingual children. *Int. J. Biling.* 13, 91–109
- Raven, J., Raven, J. C., and Court, J. H. (1998). *Standard Progressive Matrices-Parallel form*. Oxford: Oxford Psychologists Press.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Lang. Test.* 10, 355–371 doi: 10.1177/026553229301000308
- Restrepo, M. A., Morgan, G. P., and Thompson, M. S. (2013). The efficacy of a vocabulary intervention for dual-language learners with language impairment. *J. Speech Lang. Hear. Res.* 56, 748–765 doi: 10.1044/1092-4388(2012/11-0173)
- Rolstad, K., Mahoney, K., and Glass, G. V. (2005). The big picture: a meta-analysis of program effectiveness research on English language learners. *Educ. Policy* 19, 572–594 doi: 10.1177/0895904805278067
- Rose, K. (2018). *English as a Heritage Language in Israel: the Age Effect*. MA Thesis. Israel: Bar-Ilan University.
- Schoonen, R., and Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Lang. Test.* 25, 211–236 doi: 10.1177/0265532207086782
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: determinants and mechanisms. *Linguist. Approaches Biling.* 3, 3–25 doi: 10.1075/lab.3.1.01ser
- Sheng, L., Peña, E. D., Bedore, L. M., and Fiestas, C. E. (2012). Semantic deficits in Spanish-English bilingual children with language impairment. *J. Speech Lang. Hear. Res.* 55, 1–15 doi: 10.1044/1092-4388(2011/10-0254)
- Singh, L. (2014). One world, two languages: cross-language semantic priming in bilingual toddlers. *Child Dev.* 85, 755–766 doi: 10.1111/cdev.12133
- Snow, C. E. (1990). The development of definitional skill. *J. Child Lang.* 17, 697–710 doi: 10.1017/s0305000900010953
- Spencer, T. D., and Petersen, D. B. (2012). *Story Champs: a Multi-Tiered Language Intervention Program*. Laramie, WY: Language Dynamics Group.
- Spencer, T. D., Petersen, D. B., and Restrepo, M. A. (2017). *Puente de Cuentos. a Bilingual Response to Intervention Program for Preschool Children*. Laramie, WY: Language Dynamics Group.

- Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Read. Res. Q.* 21, 360–407 doi: 10.1598/rrq.21.4.1
- Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: feasible, reliable and well tolerated. *J. Telemed. Telecare* 23, 106–115 doi: 10.1177/1357633x15623921
- Unsworth, S. (2016). Early child L2 acquisition: age or input effects? neither, or both? *J. Child Lang.* 43, 608–634 doi: 10.1017/s030500091500080x
- Verhallen, M., and Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. *Appl. Linguist.* 13, 344–363 doi: 10.1093/applin/14.4.344
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to acquisition and frequency of input. *Appl. Psycholinguist.* 22, 217–234 doi: 10.1017/s0142716401002041
- Waite, M. C., Theodoros, D. G., Russell, T. G., and Cahill, L. M. (2010). Internet-based telehealth assessment of language using the CELF-4. *Lang. Speech Hear. Serv. Sch.* 41, 445–458 doi: 10.1044/0161-1461(2009/08-0131)
- Wesche, M., and Paribakht, T. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Can. Mod. Lang. Rev.* 53, 13–40 doi: 10.3138/cmlr.53.1.13
- Wiig, E. H., Secord, W. A., and Semel, E. M. (2004). *Clinical Evaluation of Language Fundamentals: Preschool 2*. San Antonio, TX: Harcourt/Psych Corporation
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Lipner, Armon-Lotem, Walters and Altman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Eye Movement Measures of Within-Language and Cross-Language Activation During Reading in Monolingual and Bilingual Children and Adults: A Focus on Neighborhood Density Effects

Veronica Whitford<sup>1\*</sup> and Marc F. Joanisse<sup>2</sup>

<sup>1</sup> Department of Psychology, University of New Brunswick, Fredericton, NB, Canada, <sup>2</sup> Department of Psychology, Brain and Mind Institute, Western University, London, ON, Canada

## OPEN ACCESS

### Edited by:

Zofia Wodniecka,  
Jagiellonian University, Poland

### Reviewed by:

Eva Commissaire,  
Université de Strasbourg, France  
Teresa Bajo,  
University of Granada, Spain  
Antonio Iniesta,  
University of Granada, Spain, in  
collaboration with reviewer TB

### \*Correspondence:

Veronica Whitford  
veronica.whitford@unb.ca

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 February 2021

**Accepted:** 27 September 2021

**Published:** 27 October 2021

### Citation:

Whitford V and Joanisse MF  
(2021) Eye Movement Measures  
of Within-Language  
and Cross-Language Activation  
During Reading in Monolingual  
and Bilingual Children and Adults:  
A Focus on Neighborhood Density  
Effects. *Front. Psychol.* 12:674007.  
doi: 10.3389/fpsyg.2021.674007

We used eye movement measures of first-language (L1) and second-language (L2) paragraph reading to investigate how the activation of multiple lexical candidates, both within and across languages, influences visual word recognition in four different age and language groups: (1) monolingual children; (2) monolingual young adults; (3) bilingual children; and (4) bilingual young adults. More specifically, we focused on within-language and cross-language orthographic neighborhood density effects, while controlling for the potentially confounding effects of orthographic neighborhood frequency. We found facilitatory within-language orthographic neighborhood density effects (i.e., words were easier to process when they had many vs. few orthographic neighbors, evidenced by shorter fixation durations) across the L1 and L2, with larger effects in children vs. adults (especially the bilingual ones) during L1 reading. Similarly, we found facilitatory cross-language neighborhood density effects across the L1 and L2, with no modulatory influence of age or language group. Taken together, our findings suggest that word recognition benefits from the simultaneous activation of visually similar word forms during naturalistic reading, with some evidence of larger effects in children and particularly those whose words may have differentially lower baseline activation levels and/or weaker links between word-related information due to divided language exposure: bilinguals.

**Keywords:** bilingualism, monolingualism, reading, eye movements, within-language and cross-language activation, orthographic neighborhood density, children, young adults

## INTRODUCTION

Though seemingly effortless, visual word recognition is a complex process that involves accessing and retrieving correct lexical representations from the mental lexicon, often among a pool of visually similar lexical candidates known as orthographic neighbors. The classic definition of orthographic neighbors includes substitution neighbors: words that resemble a target word in

all but one letter, regardless of that letter's position in the target word (Coltheart et al., 1977). For example, the English word *horse* has the following within-language (English) substitution neighbors: *horde*, *house*, *horst*, *horsy*, *Morse*, *Norse*, and *worse*. However, an updated definition also includes addition and deletion neighbors: neighbors with one additional or one fewer letter (Davis et al., 2009). For example, the English word *horse* has the following within-language (English) addition: *hoarse*, *horses*, and *horsey* and deletion: *hors* and *hose* neighbors. Alongside within-language neighbors, words can also have cross-language neighbors. For example, the English word *horse* has the following cross-language (French) substitution: *corse*, *horde*, *morse*, and *torse* and deletion: *hors* neighbors.

Whether within-language or cross-language, a target word's total number of orthographic neighbors (substitution + addition + deletion) is called its orthographic neighborhood density, and the average word frequency of its orthographic neighbors is called its orthographic neighborhood frequency. Although both properties can exert robust influences on visual word recognition, the extant research has predominantly focused on monolingual young adults, which may lack generalizability to other populations. The current study aims to address this imbalance in the literature by investigating both within-language and cross-language orthographic neighborhood density effects (while controlling for orthographic neighborhood frequency) during naturalistic paragraph reading in four participant groups differing in age and language background: (1) monolingual children; (2) monolingual young adults; (3) bilingual children; and (4) bilingual young adults. We begin with an overview of what is known about orthographic neighborhood effects among monolingual children and young adults, followed by that among bilingual children and young adults.

## Monolingual Orthographic Neighborhood Effects

### Theoretical Framework

Leading theories of monolingual visual word recognition, such as the Interactive Activation (IA) model (McClelland and Rumelhart, 1981), propose that the activation of many orthographic neighbors (i.e., a high orthographic neighborhood density) impedes lexical access of a target word. This is especially true for higher-frequency orthographic neighbors (i.e., a high orthographic neighborhood frequency), which have higher baseline activation levels and/or higher quality lexical representations (see Perfetti, 2007 for a discussion of the lexical quality hypothesis). This impedance is attributed to lateral inhibition. As orthographic units (i.e., letters and their clusters, such as bigrams) are identified, multiple lexical candidates containing these orthographic units compete for activation, especially higher-frequency candidates; they send the most lateral inhibition. Lower-frequency candidates, which have lower baseline activation levels and/or lower quality lexical representations, cannot compete as strongly for activation; they require more time to surpass the activation and overcome the lateral inhibition of their higher-frequency counterparts.

Together, these lexical candidates send negative or inhibitory feedback to the orthographic unit level, ultimately impeding the target word's lexical accessibility. Accordingly, visual word recognition is mediated by competition and inhibition from visually similar word forms.

Although the IA model predicts inhibitory orthographic neighborhood density and frequency effects, it can, however, accommodate an opposite pattern of effects—that the activation of many orthographic neighbors, including higher-frequency ones, can boost lexical access of a target word (see, for example, Andrews, 1997; Holcomb et al., 2002). The activation of multiple visually similar lexical candidates could increase the mental lexicon's overall excitation, ultimately facilitating the target word's lexical accessibility due to top-down semantic-to-lexical excitatory feedback. This facilitation could benefit lower-frequency words in particular, as they are more difficult to identify due to their lower baseline activation levels and/or lower quality lexical representations. Accordingly, visual word recognition could be mediated by facilitation (rather than competition and inhibition) from visually similar word forms.

It is important to note here that the IA model was originally developed for monolingual skilled adult readers and does not make explicit predictions regarding developmental differences in orthographic neighborhood effects. On the one hand, IA may predict larger inhibitory effects in children vs. adults. Given their younger age and developing language abilities, children's lexical representations have not benefited from as much print exposure (and language experience more generally). As a result, their words likely have differentially lower baseline activation levels and/or lower quality lexical representations, rendering them more susceptible to the effects of competition and inhibition from visually similar lexical candidates, particularly when they are higher-frequency. On the other hand, IA may also predict larger inhibitory effects in adults vs. children. Given that their lexical representations are more complex and interconnected, competition and inhibition from visually similar lexical candidates may be more pronounced due to the greater number of activated candidates.

Alternative age-related predictions are, however, possible. The activation of multiple visually similar lexical candidates, especially when lexical representations are not as entrenched in semantic memory, could accelerate the identification of orthographic units contained within target words, which, in turn, could accelerate the familiarity and overall recognition of target words. This may be particularly true for lower-frequency words, which are much less familiar to children. This would lead to larger facilitatory orthographic neighborhood density effects in children vs. adults. However, the opposite pattern may prove true: adults' activation of a greater number of lexical candidates could lead to larger facilitatory orthographic neighborhood density effects.

### Empirical Literature

Studies of monolingual orthographic neighborhood effects, which can be divided into those that have employed response-based tasks (including those with concurrent electroencephalographic/EEG recording) and those that have employed eye movement measures of reading, have

provided mixed support for the IA model. These two categories of studies are discussed in turn, with a focus on findings involving monolingual young adults, followed by those involving monolingual children.

### *Findings From Response-Based Literature*

Numerous response-based studies involving healthy monolingual young adults (aged 18–30) have reported mixed patterns of facilitatory, inhibitory, and null orthographic neighborhood density effects (e.g., Coltheart et al., 1977; Andrews, 1989, 1992; Grainger, 1992; Forster and Shen, 1996; Grainger and Jacobs, 1996; Carreiras et al., 1997; Perea and Pollatsek, 1998; Davis and Taft, 2005, Experiment 1; Pollatsek et al., 1999; Perea and Rosa, 2000, Experiment 1; Snodgrass and Mintzer, 1993; Sears et al., 1995; for reviews, see Andrews, 1997; Mathey, 2001). These between-study differences are likely driven by methodology-related factors, including the measure of orthographic neighborhood density (e.g., substitution neighbors vs. total neighbors) and the type of task used (e.g., lexical decision vs. perceptual identification tasks, which generally yield facilitatory vs. inhibitory effects, respectively, for reviews, see Andrews, 1997; Perea and Rosa, 2000; Frances et al., 2021)<sup>1</sup>. Moreover, only some studies have accounted for orthographic neighborhood frequency, namely, the presence of higher-frequency neighbors, which generally yields inhibitory effects (e.g., Grainger and Jacobs, 1996; Carreiras et al., 1997; Davis and Taft, 2005). In addition to methodology-related factors, some studies have also found that individual differences in reading and spelling abilities modulate orthographic neighborhood effects (e.g., Andrews and Hersch, 2010; Andrews and Lo, 2012), important factors that are rarely considered among monolingual, native language readers. Thus, support for IA is indeed very mixed.

Though relatively few in number, response-based studies involving healthy monolingual children (aged 7–12) have largely reported facilitatory orthographic neighborhood density effects (e.g., Laxon et al., 1988, 1994, 2002; Castles et al., 2003; Ziegler et al., 2003; Duñabeitia and Vidal-Abarca, 2008; but see Tamura et al., 2017 for lexical competition effects between newly learned words and their neighbors during a masked priming lexical decision task) (see text footnote 1). However, null orthographic neighborhood density effects can emerge when higher-frequency neighbors are accounted for (e.g., Marinus and de Jong, 2010).

Although direct comparisons with young adults' orthographic neighborhood density effects have yet to be made, there is some evidence to suggest larger effects in children. For example, some studies have found that younger children make more lexicalizations than older children and young adults when presented with non-words that resemble words, such as *cholate* vs. *chocolate* (e.g., Sebastián-Gallés and Parraño, 1995; Perea and Fraga, 2006; Perea and Estévez, 2008). Such effects can be attributed to some developing readers' holistic strategies that

rely more on coarse-grained orthographic codes (as opposed to fine-grained ones) during visual word recognition. More specifically, holistic strategies lean on minimal orthographic units needed to convey word identity, regardless of exact letter ordering (see Grainger and Ziegler, 2011)—strategies that may ultimately contribute to larger facilitatory orthographic neighborhood effects in children compared to adults. Though such a pattern of findings would refute IA's original predictions, it can, however, be explained through the alternative interpretation of the model (discussed previously). We note, however, that children's use of such strategies may vary as a function of the orthographic transparency of their known languages, with a potentially greater use when reading in opaque languages due to inconsistent grapheme-to-phoneme correspondences (e.g., Ziegler and Goswami, 2006; Rau et al., 2015, 2016).

### *Findings From Eye Movement Literature*

Surprisingly, few monolingual studies have employed eye movement measures of reading to examine orthographic neighborhood effects, despite having several advantages over response-based tasks. These include contextualized stimuli, such as sentences and passages, instead of isolated words; naturalistic or ecologically valid tasks, such as reading for comprehension, instead of making artificial decisions to target words; and greater temporal sensitivity—being able to examine both early and late stages of word processing through different measures, such as gaze duration (i.e., the sum of all fixation durations on a word during the first pass, reflecting lexical access) and total reading time (i.e., the grand sum of all fixation durations on a word, reflecting post-lexical integration), instead of measuring global reaction times and accuracy scores for target words (Rayner, 1998, 2009). As a result, response-based tasks and eye movement measures of reading probe fundamentally different language processes (see Kuperman et al., 2013).

To date, all eye movement studies have focused on healthy monolingual young adults, and only one has examined orthographic neighborhood density effects during sentence reading (Pollatsek et al., 1999, Experiment 2). It found inhibitory effects during both early stage (gaze duration) and late-stage (total reading time) word processing. However, when higher-frequency neighbors were accounted for, the effects were facilitatory during early stage word processing (skipping rate—i.e., the probability of fixating a word during the first pass), but inhibitory during late-stage word processing (regressions out—i.e., backward eye movements to a word indicative of rereading). This suggests that the activation of multiple lexical candidates may have led participants to misread or misidentify words on the first pass, as facilitatory effects during lexical access were followed by inhibitory effects during post-lexical integration. Other eye movement studies have also reported inhibitory effects of higher-frequency neighbors (Perea and Pollatsek, 1998; Slattery, 2009; Gregg and Inhoff, 2016; but see Sears et al., 2006 for null effects). Accordingly, it appears that monolingual orthographic neighborhood effects are largely inhibitory during sentence reading—findings that support IA. However, the nature of these effects among monolingual children is currently unknown. The

<sup>1</sup>The use of masked priming paradigms has also contributed to different patterns of findings, including monolingual orthographic priming studies that have manipulated neighborhood density and/or frequency in adults (e.g., Davis and Lupker, 2006; Nakayama et al., 2008; Massol et al., 2010) and in children (e.g., Castles et al., 2003; Tamura et al., 2017).

current study, which includes both monolingual age groups, will fill this crucial gap in the empirical literature.

## Bilingual Orthographic Neighborhood Effects

### Theoretical Framework

Bilingualism has important consequences for how word forms are represented and retrieved from the mental lexicon during first-language (L1) and second-language (L2) visual word recognition. One of these consequences is the automatic, non-selective activation of both target and non-target language lexical representations—a phenomenon called cross-language activation. In other words, even in unilingual language contexts, bilinguals must access and retrieve correct lexical representations from a pool of visually similar lexical candidates across their known languages.

Leading theories of bilingual visual word recognition, such as the Bilingual Interactive Activation (BIA; Dijkstra and van Heuven, 1998) and Bilingual Interactive Activation Plus (BIA+; Dijkstra and van Heuven, 2002) models, which are bilingual adaptations of the monolingual IA model, propose that bilinguals have an integrated lexicon, wherein both their languages are represented. As a result, when bilinguals are visually presented with a word, similar lexical candidates from their known languages are coactivated due to spreading activation during bottom-up processing (e.g., identification of orthographic units). The activation of these lexical candidates can facilitate or inhibit word recognition, depending on a variety of factors. These include methodology-related factors, such as the nature of the task (e.g., making word judgments vs. reading for comprehension), nature of the cross-linguistic overlap (e.g., cognates, interlingual homographs, or cross-language orthographic neighbors), degree of contextual constraint (e.g., isolated words vs. sentences or paragraphs), and global language context (e.g., instructions and/or stimuli presented in the L1, L2, or both). These also include participant-related factors, such as age and manner of L1/L2 acquisition, L1/L2 dominance, L1/L2 proficiency, and domain-general executive control abilities (for reviews, see Van Assche et al., 2012; Kroll et al., 2016; Titone et al., 2016; Whitford et al., 2016; Lauro and Schwartz, 2017).

Regarding orthographic neighborhood effects, the BIA/BIA+ models make similar predictions as those of the IA model. The activation of many orthographic neighbors both within and across languages impedes lexical access of a target word. This is especially true for higher-frequency L1 orthographic neighbors, which have higher baseline activation levels and/or stronger links between different types of word-related information, such as orthography, phonology, and semantics (see Gollan et al., 2008, 2011 for a discussion of the weaker links hypothesis—a bilingual adaptation of Perfetti's lexical quality hypothesis). Again, this impedance is attributed to lateral inhibition, which is heightened for lower-frequency L2 words; they have not benefited from as much experience, resulting in lower baseline activation levels and/or weaker links between word-related information, and, ultimately, a reduced

capacity to compete with activation of their higher-frequency L1 counterparts (which send the most lateral inhibition). Although these models predict inhibitory orthographic neighborhood effects, they can, however, accommodate facilitatory ones. The activation of many orthographic neighbors (both within and across languages) could increase the mental lexicon's overall excitation and facilitate a target word's accessibility due to top-down semantic-to-lexical excitatory feedback, especially when it is more difficult to identify, as is the case with lower-frequency L2 words.

Like the IA model, the BIA/BIA+ models also do not make explicit predictions regarding developmental differences in orthographic neighborhood effects. However, their predictions would likely be similar. One possibility is larger inhibitory effects in children vs. adults, especially for lower-frequency L2 words. Again, this may be driven by children's reduced age, print exposure, and language abilities, particularly in their weaker language: L2. Another possibility is larger inhibitory effects in adults vs. children, again, due to the greater competition and inhibition that ensues when more lexical candidates are activated. These factors could, however, contribute to larger facilitatory effects in these age groups, respectively. Regarding larger facilitatory effects in children, lexical representations that are not as entrenched in the mental lexicon, as is the case for lower-frequency L2 words among developing readers, could differentially benefit from the activation of multiple orthographic neighbors, thereby boosting their lexical accessibility. Regarding larger facilitatory effects in adults, the activation of a greater number of lexical candidates could boost overall lexical accessibility, which would particularly benefit the recognition of lower-frequency L2 words.

### Empirical Literature

Studies of bilingual orthographic neighborhood effects, which can also be divided into those that have employed response-based tasks (including those with concurrent EEG recording) and those that have employed eye movement measures of reading, have provided mixed support for the BIA/BIA+ models. These two categories of studies are discussed in turn, with a focus on findings involving bilingual young adults. Although the current study represents the first investigation of orthographic neighborhood effects in bilingual children, relevant findings from studies investigating other aspects of orthographic processing are discussed.

#### *Findings From Response-Based Literature*

The bilingual literature parallels the monolingual literature; it has reported mixed patterns of facilitatory, inhibitory, and null within-language and cross-language orthographic neighborhood density effects among healthy bilingual young adults (aged 18–30) across their L1 and L2 (e.g., Beauvillain, 1992; de Groot et al., 2002; Dirix et al., 2017; Commissaire et al., 2019, Experiment 1; Grainger and Dijkstra, 1992; Van Heuven et al., 1998; Lemhöfer et al., 2008; Midgley et al., 2008; Grossi et al., 2012; Meade et al., 2018; Mulder et al., 2018). As discussed previously, these between-study differences are likely driven by methodology-related factors, including whether orthographic neighborhood

frequency was accounted for, as well as by participant-related factors. Thus, support for BIA/BIA+ is indeed very mixed.

Although no prior response-based studies have investigated orthographic neighborhood effects in bilingual children, there is evidence that their visual word recognition is differentially mediated by cross-language orthographic overlap, especially when they are younger in age (for a review, see van Hell, 2020). For instance, Duñabeitia et al. (2016) tested a large sample ( $N = 100$ ) of balanced Spanish-Basque bilingual children (aged 8–15) on L1 and L2 translation recognition tasks and found that the younger children's performance was more sensitive to the target words' cognate status, with greater orthographic overlap facilitating word recognition (see also Schröter and Schroeder, 2016, 2018 and Duñabeitia et al., 2020, for similar findings involving similar and other aspects of cross-language orthographic processing in children during L1 and L2 lexical decision tasks). These effects have been attributed to younger children's reduced print exposure and developing language control systems, which render their lexical representations more susceptible to the effects of cross-language activation. Based on this work, larger facilitatory orthographic neighborhood effects in bilingual children vs. adults are likely. Such a pattern of findings would support the alternative interpretation of BIA/BIA+ (discussed previously).

### Findings From Eye Movement Literature

Though less than a handful, eye movement studies of reading have generally reported facilitatory within-language and cross-language orthographic neighborhood density effects among healthy bilingual young adults (aged 18–30) across their L1 and L2. Thus, these findings refute BIA/BIA+'s original predictions and suggest that the models may require modifications to account for certain aspects of within-language and cross-language activation during natural reading. Indeed, these findings support the models' alternative interpretation.

In the earliest of studies, Whitford et al. (2016) found facilitatory cross-language orthographic neighborhood density effects during L1 and L2 paragraph reading in a large sample ( $N = 117$ ) of balanced English-French bilingual young adults. Words with many vs. fewer cross-language neighbors were easier to process, evidenced by shorter gaze durations and total reading times. This was especially true for lower-frequency L2 words, which have differentially lower baseline activation levels and/or weaker links between word-related information. The patterns of within-language orthographic neighborhood density effects differed across the L1 and L2; they were null vs. facilitatory. The study did, however, have two important limitations that were addressed in subsequent work: it only included substitution neighbors and did not account for orthographic neighborhood frequency.

Consistent with Whitford et al.' (2016) study, Dirix et al. (2017, Experiment 2) also found largely facilitatory cross-language orthographic neighborhood density effects during L1 and L2 novel reading in a small sample ( $N = 19$ ) of unbalanced Dutch-English bilingual young adults. Although the effects were rather limited during L1 reading, words with many vs. fewer cross-language orthographic neighbors were, again, processed

more easily, evidenced by shorter gaze durations and total reading times. The patterns of within-language neighborhood density effects differed across the L1 and L2. In the L1, they were facilitatory for lower-frequency words and inhibitory for higher-frequency words, whereas in the L2, they were entirely facilitatory.

Extending their previous study to examine age differences in orthographic neighborhood density effects, Whitford and Titone (2019) found facilitatory within-language and cross-language neighborhood density effects during L1 and L2 paragraph reading in large samples ( $n = 62$  each) of balanced French-English bilingual younger and older adults (aged 60+), matched on gender, education, L1/L2 background, and L1/L2 proficiency (both objective and subjective). Although their findings patterned with those of their earlier study, larger effects were observed among older adults. Thus, despite having benefited from more life-long print exposure (and language experience more generally), older adults' lexical accessibility may be negatively mediated by age-related changes in cognitive and sensory processing.

Although no prior eye movement reading studies have investigated orthographic neighborhood effects in bilingual children, a recent study suggests that their visual word recognition may be positively mediated by cross-language orthographic overlap. Bosma and Nota (2020) found that a group ( $N = 37$ ) of L2-dominant Frisian-Dutch bilingual children (aged 9–12) were sensitive to target words' cognate status, with greater orthographic overlap (form-identical, followed by form-non-identical cognates) facilitating word recognition, evidenced by shorter gaze durations and total reading times.

Taken together, the above-reviewed bodies of literature suggest that visual word recognition is influenced by orthographic neighborhood effects in various ways across different experimental tasks and participant groups. Here, we clarify and unify these distinct bodies of literature by examining how both within-language and cross-language orthographic neighborhood effects influence visual word recognition during naturalistic reading in different age groups (children, adults) and language groups (monolinguals, bilinguals), and whether the observed findings can be captured by the IA and BIA/BIA+ models. Thus, this work will further our understanding of a relatively understudied potential moderator of within-language and cross-language activation in diverse groups of people: orthographic neighborhood density.

## The Current Study

We investigated how monolingual and bilingual children's and young adults' L1 and L2 eye movement reading behavior was influenced by orthographic neighborhood density (both cross-language and within-language, where applicable), while controlling for the presence of higher-frequency orthographic neighbors. Based on previous findings from the bilingual eye movement reading literature (which are largely consistent with BIA/BIA+'s alternative explanations), we predicted facilitatory cross-language and within-language orthographic neighborhood density effects across both early and late reading stages: words with higher orthographic neighborhood densities should be

easier to process, evidenced by shorter gaze durations and total reading times. However, we also predicted modulatory effects of age group and language group based on the lower word baseline activation levels and/or weaker links that some of the participant groups may experience (namely, bilinguals and children). Thus, our specific hypotheses were as follows:

- (1) During L1 reading, larger facilitatory within-language (L1) effects in bilinguals vs. monolinguals, as well as in children vs. adults.
- (2) During L2 reading, larger facilitatory within-language (L2) effects in bilingual children vs. bilingual adults.
- (3) During L1 reading, larger cross-language (L2) effects in bilingual children vs. bilingual adults.
- (4) During L2 reading, larger cross-language (L1) effects in bilingual children vs. bilingual adults.

## MATERIALS AND METHODS

### Participants

Participants were the same as those included in Whitford and Joanisse (2018). They comprised four groups: (1) English monolingual children aged 7–12 ( $n = 34$ ); (2) English-French bilingual children aged 7–12 ( $n = 33$ ); (3) English monolingual adults aged 18–21 ( $n = 30$ ); and (4) English-French bilingual adults aged 18–21 ( $n = 30$ ). The children were recruited from English-language, French-language, and French immersion elementary schools in London, Ontario, Canada, and the adults were recruited from Western University (most of the bilingual adults attended French immersion schools as children). All participants had English as their first acquired and dominant language (L1), and all bilingual participants had French as their second acquired and weaker language (L2). Note that some of the monolingual children and adults had some French instruction through the Ontario educational curriculum; however, all self-identified as functionally monolingual. All participants were typically developing, with no uncorrected visual or hearing impairments, and no language, learning, neurological, or psychiatric disorders. The study was part of a larger experimental protocol that lasted 3 hours. Participants received a \$30 movie gift card or course credit as compensation. The study was approved by Western University's Non-Medical Research Ethics Board (106319/106601).

Participants completed three background measures. First, adaptations of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) were used to assess participants' demographic and language backgrounds, including L1/L2 age of acquisition (AoA) and current L1/L2 exposure. Second, the Word Reading and Pseudoword Decoding subtests of the Wechsler Individual Achievement Test—Second Edition (WIAT-II; English-Canadian and French-Canadian adaptations; Wechsler, 2005) were used to assess participants' L1 and L2 word-level reading skills. More specifically, the Word Reading subtest measured accuracy of word recognition without contextual clues and the Pseudoword Decoding subtest measured accuracy of deciphering non-sense words. Participants read aloud a list of

words (maximum: 131) and a list of made-up words (maximum: 55) that increased in difficulty. Raw subtest scores were converted to age-based standard scores ( $M = 100 \pm 15$ ). Third, the Test of Non-Verbal Intelligence—Third Edition (TONI-III; Brown et al., 1997) was used to assess participants' non-verbal IQ. More specifically, participants completed sequences of shapes (maximum: 45) by selecting one option among six possible response options. Raw scores were converted to age-based standard scores ( $M = 100 \pm 15$ ).

Participant characteristics are presented in **Tables 1, 2**, which demonstrate that the two groups of children and the two groups of adults were matched on age, sex, education, parental socioeconomic status (SES) based on the Hollingshead Occupational Scale (Hollingshead, 1975), non-verbal IQ, self-report (i.e., LEAP-Q) measures of L1 history and proficiency, and objective (i.e., WIAT-II) measures of L1 reading ability

**TABLE 1 |** Characteristics of the child participant groups.

|                                                              | Monolingual children<br>( $n = 34$ )<br>[mean (SD)] | Bilingual children<br>( $n = 33$ )<br>[mean (SD)] |
|--------------------------------------------------------------|-----------------------------------------------------|---------------------------------------------------|
| Age (years)                                                  | 9.82 (1.10)                                         | 10.02 (1.32)                                      |
| Sex (male:female ratio)                                      | 14:20                                               | 13:20                                             |
| Education (years)                                            | 4.09 (1.08)                                         | 4.21 (1.39)                                       |
| Parental SES <sup>a</sup>                                    | 3.00 (1.18)                                         | 2.88 (1.36)                                       |
| TONI-III (standard scores)                                   | 109.88 (17.10)                                      | 117.18 (18.04)                                    |
| <b>AoA; Age of fluency (years)</b>                           |                                                     |                                                   |
| L1                                                           | Birth (–); 2.71 (0.95)                              | Birth (–); 2.43 (1.17)                            |
| L2***                                                        | 7.42 (1.82); Never (–)                              | 3.82 (1.66); 5.57 (1.96)                          |
| <b>Reading AoA; Age of reading fluency (years)</b>           |                                                     |                                                   |
| L1                                                           | 4.35 (0.96); 6.05 (0.95)                            | 4.48 (1.14); 6.23 (1.37)                          |
| L2***                                                        | 8.28 (1.07); Never (–)                              | 5.47 (1.05); 7.36 (1.44)                          |
| <b>Current language exposure (% time)</b>                    |                                                     |                                                   |
| L1***                                                        | 95.53 (5.66)                                        | 58.03 (12.93)                                     |
| L2***                                                        | 4.47 (5.66)                                         | 39.70 (13.11)                                     |
| <b>Current reading exposure (% time)</b>                     |                                                     |                                                   |
| L1***                                                        | 99.79 (0.88)                                        | 65.30 (25.98)                                     |
| L2***                                                        | 0.21 (0.88)                                         | 33.58 (25.35)                                     |
| <b>L1 self-report proficiency measures (1–7)<sup>b</sup></b> |                                                     |                                                   |
| Reading ability                                              | 6.06 (1.41)                                         | 5.64 (1.41)                                       |
| Overall competence                                           | 6.15 (1.31)                                         | 5.88 (1.11)                                       |
| <b>L2 self-report proficiency measures (1–7)<sup>b</sup></b> |                                                     |                                                   |
| Reading ability***                                           | 1.06 (0.24)                                         | 4.58 (1.28)                                       |
| Overall competence***                                        | 1.06 (0.24)                                         | 4.67 (1.31)                                       |
| <b>L1 WIAT-II (standard scores)</b>                          |                                                     |                                                   |
| Word Reading                                                 | 99.44 (12.58)                                       | 99.15 (17.38)                                     |
| Pseudoword Decoding                                          | 106.26 (15.61)                                      | 103.12 (17.22)                                    |
| <b>L2 WIAT-II (standard scores)</b>                          |                                                     |                                                   |
| Word Reading                                                 | –                                                   | 88.55 (23.77)                                     |
| Pseudoword Decoding                                          | –                                                   | 95.70 (20.73)                                     |

SES, socioeconomic status; TONI-III, Test of Non-Verbal Intelligence—3rd Edition; AoA, age of acquisition; L1, first-language; L2, second-language; WIAT-II, Wechsler Individual Achievement Test—2nd Edition.

<sup>a</sup>Scale from 1 (major professional) to 9 (unemployed).

<sup>b</sup>Scale from 1 (beginner) to 7 (native-like).

\*\*\* $p < 0.001$ .

**TABLE 2 |** Characteristics of the adult participant groups.

|                                                              | Monolingual adults<br>( <i>n</i> = 30)<br>[mean ( <i>SD</i> )] | Bilingual adults<br>( <i>n</i> = 30)<br>[mean ( <i>SD</i> )] |
|--------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------|
| Age (years)                                                  | 18.67 (0.94)                                                   | 18.33 (0.60)                                                 |
| Sex (male:female ratio)                                      | 10:20                                                          | 5:25                                                         |
| Education (years)                                            | 13.35 (0.52)                                                   | 13.17 (0.30)                                                 |
| Parental SES <sup>a</sup>                                    | 2.63 (1.14)                                                    | 2.27 (1.09)                                                  |
| TONI-III (standard scores)                                   | 99.60 (11.84)                                                  | 99.39 (14.04)                                                |
| <b>AoA; Age of fluency (years)</b>                           |                                                                |                                                              |
| L1                                                           | Birth (–); 3.72 (1.80)                                         | Birth (–); 3.68 (1.75)                                       |
| L2***                                                        | 8.96 (2.46); Never (–)                                         | 5.53 (2.42); 10.95 (4.51)                                    |
| <b>Reading AoA; Age of reading fluency (years)</b>           |                                                                |                                                              |
| L1                                                           | 4.52 (1.32); 6.50 (1.96)                                       | 5.03 (1.49); 7.07 (1.57)                                     |
| L2***                                                        | 9.90 (2.24); Never (–)                                         | 7.13 (2.08); 11.10 (3.56)                                    |
| <b>Current language exposure (% time)</b>                    |                                                                |                                                              |
| L1**                                                         | 99.70 (0.97)                                                   | 86.41 (19.71)                                                |
| L2***                                                        | 0.30 (0.97)                                                    | 12.73 (19.67)                                                |
| <b>Current reading exposure (% time)</b>                     |                                                                |                                                              |
| L1***                                                        | 100.00 (0.00)                                                  | 86.34 (18.84)                                                |
| L2***                                                        | 0.00 (0.00)                                                    | 13.66 (18.84)                                                |
| <b>L1 self-report proficiency measures (1–7)<sup>b</sup></b> |                                                                |                                                              |
| Reading ability                                              | 6.83 (0.73)                                                    | 6.67 (0.74)                                                  |
| Overall competence                                           | 6.93 (0.36)                                                    | 6.70 (0.53)                                                  |
| <b>L2 self-report proficiency measures (1–7)<sup>b</sup></b> |                                                                |                                                              |
| Reading ability***                                           | 1.43 (0.62)                                                    | 5.20 (0.98)                                                  |
| Overall competence***                                        | 1.17 (0.37)                                                    | 4.83 (1.07)                                                  |
| <b>L1 WIAT-II (standard scores)</b>                          |                                                                |                                                              |
| Word Reading                                                 | 111.80 (6.55)                                                  | 112.43 (5.39)                                                |
| Pseudoword Decoding                                          | 105.73 (11.85)                                                 | 109.07 (8.12)                                                |
| <b>L2 WIAT-II (standard scores)</b>                          |                                                                |                                                              |
| Word Reading                                                 | –                                                              | 81.18 (18.92)                                                |
| Pseudoword Decoding                                          | –                                                              | 97.70 (12.78)                                                |

SES, socioeconomic status; TONI-III, Test of Non-Verbal Intelligence—3rd Edition; AoA, age of acquisition; L1, first-language; L2, second-language; WIAT-II, Wechsler Individual Achievement Test—2nd Edition.

<sup>a</sup>Scale from 1 (major professional) to 9 (unemployed).

<sup>b</sup>Scale from 1 (beginner) to 7 (native-like).

\*\**p* < 0.01; \*\*\**p* < 0.001.

(all *p*-values > 0.5). Expectedly, both monolingual groups had significantly lower L2 proficiency than their bilingual counterparts based on self-report measures (all *p*-values < 0.001); the monolingual groups (especially the children) lacked the proficiency needed to complete the objective measures. Both bilingual groups had significantly lower L2 vs. L1 proficiency based on both self-report and objective measures (all *p*-values < 0.05). Although the two groups of children were matched as closely as possible to their adult counterparts, including on parental SES and WIAT-II Pseudoword Decoding (across their known languages), they significantly differed on a number of other measures, including non-verbal IQ, L1 WIAT-II Word Reading, and current L1/L2 exposure (for the bilingual groups) (all *p*-values < 0.01). Thus, we decided to control for these differences in our analyses. Nonetheless, the pattern of results (reported subsequently) remained unchanged even when subsets of adults matched

even more closely to the groups of children were included in the analyses.

## Materials

Stimuli were the same as those in Whitford and Joanisse (2018): English and French versions of four paragraphs (two fiction and two non-fiction short stories), drawn from the Reading Comprehension subtest of the WIAT-II (English-Canadian and French-Canadian adaptations; Wechsler, 2005). The paragraphs were representative of those read in elementary educational settings across Canada; thus, they had a high degree of ecological validity. The English and French versions of the paragraphs contained a comparable number of words (105, 87, 103, and 195 words vs. 118, 95, 109, and 200 words). Important lexical characteristics were obtained for the words of each paragraph, including length, frequency, predictability, mean bigram frequency (both within-language and cross-language), total phonological neighborhood density (both within-language and cross-language), total orthographic neighborhood density (both within-language and cross-language), orthographic neighborhood density of higher-frequency neighbors (both within-language and cross-language), and orthographic neighborhood frequency of higher-frequency neighbors (both within-language and cross-language).

The paragraphs' English and French lexical characteristics were obtained as follows. Subtitle word frequency values (in occurrences per million words) were gathered from SUBTLEX-US (Brysbaert and New, 2009) via the English Lexicon Project (Balota et al., 2007) and Lexique (New et al., 2001), respectively. Mean bigram frequencies were computed by dividing summated bigram frequencies from WordGen (Duyck et al., 2004) by word length (following Dirix et al., 2017). Word predictability values were derived through computerized cumulative cloze tasks involving separate samples of native English (*n* = 30) and native French (*n* = 30) young adult participants, who guessed the words of each paragraph one at a time, until the entire paragraph was presented on the computer screen (following Miellet et al., 2007; Whitford and Titone, 2012, 2014, 2017, 2019). Neighborhood density and neighborhood frequency values were gathered from the Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities (CLEARPOND; Marian et al., 2012). Paragraph characteristics are presented in **Supplementary Table A1**.

A total of 210 language-unique target words were selected from the paragraphs and included in the analyses. Exclusions were as follows: line-initial, line-final, function, proper noun, punctuated, and repeated words, as well as cognates and interlingual homographs (Pollatsek et al., 2006; Miellet et al., 2007; Whitford and Titone, 2012, 2014, 2017, 2019). Target word characteristics are presented in **Supplementary Table A2**.

## Apparatus

Right eye movements were sampled at 1 kHz using an EyeLink 1000 desktop-mounted eye-tracker (SR-Research, Ontario, Canada). The paragraphs were viewed binocularly on a 21" ViewSonic CRT monitor (screen resolution: 1,024 × 768 pixels; viewing distance: 60 cm). Depending on their length,

the paragraphs were presented on one or two black display pages in yellow 14-point Courier New font using Experiment Builder (SR-Research, Ontario, Canada). Each display page had a maximum of 10 lines of text, 70 characters per line, and 2 characters per 1° of visual angle. Eye movements were calibrated with a nine-point grid (average fixation error:  $< 0.5^\circ$  of visual angle following validation). A padded head-rest minimized head movements during reading.

## Procedure

The procedure was the same as that of Whitford and Joanisse (2018). After providing both oral and written assent and/or consent, participants read the four paragraphs (two in their L1 and two in their L2) silently and naturally for comprehension while their eye movements were monitored. Paragraph version (1, 2, 3, 4) and paragraph language (L1, L2) were counterbalanced across participants. Calibration procedures were performed before each paragraph was read. Comprehension was assessed via four open-ended, orally administered questions after reading each paragraph (total score: 16). Correct, partially correct, and incorrect answers were scored as 1, 0.5, and 0, respectively (Radach et al., 2008; Whitford and Titone, 2012, 2014, 2017, 2019). Subsequently, participants (or the caregivers/parents of children) completed the LEAP-Q, followed by the WIAT-II Word Reading and Pseudoword Decoding subtests (counterbalanced across participants) in English (if monolingual) or in both English and French (if bilingual), and the TONI-III.

## RESULTS

### Reading Comprehension Performance

A one-way ANOVA revealed comparable L1 (English) reading comprehension accuracy between the four participant groups [ $F_{(3, 123)} = 0.47, p = 0.703$ ]. Moreover, a two-way ANOVA revealed comparable L1 (English) and L2 (French) reading comprehension accuracy between the two bilingual groups [ $F_{(1, 122)} = 0.05, p = 0.822$ ]. Thus, there were no between-group or between-language differences in reading comprehension performance. Means and standard deviations are presented in Table 3.

### Eye Movement Reading Performance

The EyeLink 1000 software identified fixations (pauses) and saccades (eye movements), which had a minimum velocity of  $30^\circ/\text{s}$ , minimum acceleration of  $8,000^\circ/\text{s}^2$ , and minimum change

in eye position of  $0.15^\circ$ . A lower cut-off of 80 ms was applied to all fixations ( $< 5\%$  of data); however, an upper cut-off was not applied to maximize data inclusion (maximum fixation duration: 2,605 ms made by a bilingual child reading in their L2).

We examined two eye movement measures. One reflected early stage reading (i.e., lexical access): gaze duration (i.e., the sum of all fixation durations on a word during the first pass). One reflected late-stage reading (i.e., post-lexical integration): total reading time (i.e., the grand sum of all fixation durations on a word). Only fixations on the 210 language-unique target words were included in the analyses.

We used linear mixed-effects models (LMMs) to analyze the data via the lme4 package (Bates, 2007; Bates et al., 2015) in version 4.0.4 of R (Baayen, 2008; R Development Core Team, 2021). We ran four models across the two eye movement measures. They examined between-group differences in: (1) within-language (L1) neighborhood density effects during L1 reading; (2) within-language (L2) neighborhood density effects during L2 reading; (3) cross-language (L2) neighborhood density effects during L1 reading; and (4) cross-language (L1) neighborhood density effects during L2 reading. The fixed effects (i.e., factors of theoretical interest), control predictors (i.e., covariates), and random effects (i.e., random intercepts and/or slopes for participants and items) for each model are reported subsequently. Across all models, categorical variables were deviation coded ( $-0.5, 0.5$ ), where the mean of each level was compared to the grand mean, and continuous variables were scaled (i.e., standardized, z-scored) to reduce collinearity. Of note, only significant effects (i.e., those with  $|t|$  values  $> 1.96$ , corresponding to  $\alpha = 0.05$ ) involving the fixed effects and their interactions are reported subsequently; however, complete model outputs can be found in **Supplementary Appendix**.

### Model 1: Within-Language (L1) Neighborhood Density Effects on L1 Reading

All four participant groups were included in this analysis. The fixed factors were age group (children vs. adults), language group (monolingual vs. bilingual), and total within-language (L1) orthographic neighborhood density (continuous). The word-related control predictors were length (continuous), frequency (continuous, log-transformed), predictability (continuous), mean within-language bigram frequency (continuous, log-transformed), total within-language phonological neighborhood density (continuous), within-language orthographic neighborhood density of higher-frequency neighbors (continuous), and within-language orthographic neighborhood frequency of higher-frequency neighbors (continuous). The participant-related control predictors were L1 WIAT-II Word Reading standard scores (continuous), current L1 exposure (continuous), and TONI-III standard scores (continuous). The random effects were random intercepts for participants and paragraph version (following Whitford and Titone, 2012, 2014, 2017, 2019). Complete model outputs for this analysis can be found in **Supplementary Table A3**.

The effect of age group was significant for gaze duration ( $\beta = 62.69, SE = 18.08, t = 3.47, p = 0.001$ ). Children had

**TABLE 3 |** Paragraph reading comprehension performance (% correct).

|    | Monolingual<br>children<br>( <i>n</i> = 34)<br>[mean (SD)] | Bilingual<br>children<br>( <i>n</i> = 33)<br>[mean (SD)] | Monolingual<br>adults<br>( <i>n</i> = 30)<br>[mean (SD)] | Bilingual<br>adults<br>( <i>n</i> = 30)<br>[mean (SD)] |
|----|------------------------------------------------------------|----------------------------------------------------------|----------------------------------------------------------|--------------------------------------------------------|
| L1 | 84.01 (11.74)                                              | 82.46 (17.54)                                            | 86.25 (11.42)                                            | 87.04 (13.10)                                          |
| L2 | —                                                          | 77.83 (18.66)                                            | —                                                        | 81.47 (14.77)                                          |

L1, first-language; L2, second-language.

longer gaze durations (343 vs. 251 ms) than adults, reflecting more effortful reading. The effect of total (L1) orthographic neighborhood density was near-significant for total reading time ( $\beta = -15.64$ ,  $SE = 8.22$ ,  $t = -1.90$ ,  $p = 0.057$ ). Words with many vs. few within-language neighbors were easier to process, evidenced by shorter total reading times (378 vs. 431 ms).<sup>2</sup>

Moreover, the three-way interaction between age group, language group, and total within-language (L1) orthographic neighborhood density was significant for total reading time ( $\beta = 41.36$ ,  $SE = 19.46$ ,  $t = 2.13$ ,  $p = 0.034$ ). To facilitate interpretation of the higher-order interaction, we ran separate follow-up models with either monolingual adults, bilingual adults, or monolingual children as the baseline. Significantly or marginally larger facilitatory neighborhood density effects were found between the following groups: monolingual children vs. monolingual adults ( $\beta = -16.86$ ,  $SE = 8.65$ ,  $t = -1.95$ ,  $p = 0.052$ ); bilingual children vs. monolingual adults ( $\beta = -38.88$ ,  $SE = 13.60$ ,  $t = -2.86$ ,  $p = 0.004$ ); monolingual children vs. bilingual adults ( $\beta = -34.08$ ,  $SE = 13.92$ ,  $t = -2.45$ ,  $p = 0.014$ ); bilingual children vs. bilingual adults ( $\beta = -57.16$ ,  $SE = 15.97$ ,  $t = -3.58$ ,  $p < 0.001$ ); and bilingual children vs. monolingual children ( $\beta = -23.08$ ,  $SE = 13.25$ ,  $t = -1.74$ ,  $p = 0.082$ ). No significant difference was found between the adult groups ( $\beta = -18.27$ ,  $SE = 14.26$ ,  $t = -1.28$ ,  $p = 0.200$ ). Thus, as can be seen in **Figure 1**, both groups of children exhibited larger facilitatory neighborhood density effects than both groups of adults; words were easier to process when they had many vs. few within-language neighbors, evidenced by shorter total reading times (see text footnote 2). However, the magnitude of these effects was most pronounced in bilingual children; they found words

with few within-language neighbors especially difficult to process, evidenced by differentially longer total reading times.

### Summary of Model 1's Findings

We observed larger facilitatory total within-language (L1) orthographic neighborhood density effects in children vs. adults during late-stage reading, especially among bilingual children.

### Model 2: Within-Language (L2) Neighborhood Density Effects on L2 Reading

Only the bilingual groups were included in this analysis. The fixed factors were age group (children vs. adults) and total within-language (L2) orthographic neighborhood density (continuous). The word-related control predictors were the same as in the previous model. The participant-related control predictors were L2 WIAT-II Word Reading standard scores (continuous), current L2 exposure (continuous), and TONI-III standard scores (continuous). The random effects were the same as in the previous model. Complete model outputs for this analysis can be found in **Supplementary Table A4**.

Although no effects reached significance, the effect of total within-language (L2) orthographic neighborhood density was near-significant for total reading time ( $\beta = -32.95$ ,  $SE = 17.30$ ,  $t = -1.90$ ,  $p = 0.057$ ). Words with many vs. few within-language neighbors were easier to process, evidenced by numerically shorter total reading times (491 vs. 679 ms).<sup>2</sup>

### Summary of Model 2's Findings

We observed numerically facilitatory total within-language (L2) orthographic neighborhood density effects during late-stage reading that were age-invariant.

### Model 3: Cross-Language (L2) Neighborhood Density Effects on L1 Reading

All four participant groups were included in this analysis. The fixed factors were the same as those in Model 1, except

<sup>2</sup> Although total orthographic neighborhood density was analyzed continuously in all models, it was dichotomized using a median split for the presentation of means to facilitate comparisons with standard analyses of variance (ANOVAs).

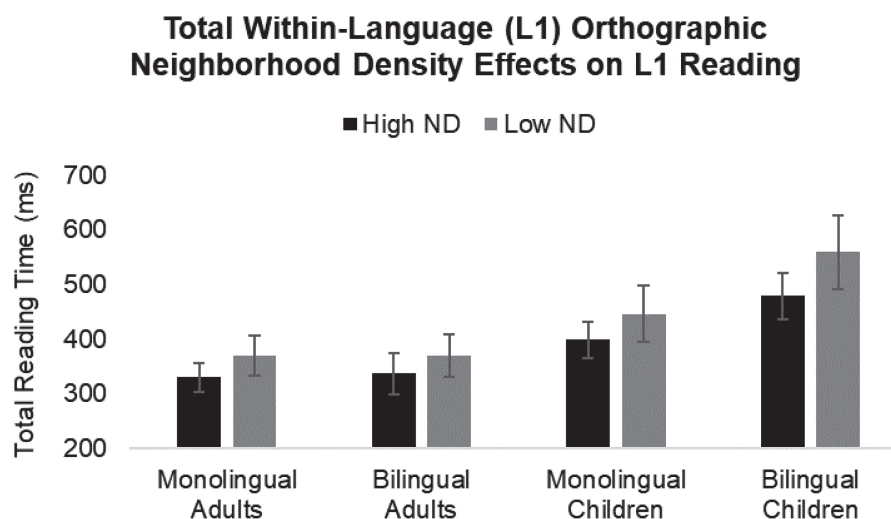


**FIGURE 1 |** The effect of total within-language (L1) orthographic neighborhood density on the monolingual and bilingual age groups' L1 total reading times. Means are plotted. Error bars represent  $\pm 1$  standard error of the mean.

that total cross-language (L2) orthographic neighborhood density (continuous) was included instead. The word-related and participant-related control predictors were also the same as those in Model 1, with the addition of mean cross-language bigram frequency (continuous, log-transformed), total cross-language phonological neighborhood density (continuous), cross-language orthographic neighborhood density of higher-frequency neighbors (continuous), and cross-language orthographic neighborhood frequency of higher-frequency neighbors (continuous). The random effects were the same as those in previous models. Complete model outputs for this analysis can be found in **Supplementary Table A5**.

The effect of age group was significant for gaze duration ( $\beta = 45.67$ ,  $SE = 21.72$ ,  $t = 2.10$ ,  $p = 0.037$ ). Children had longer gaze durations than adults (343 vs. 251 ms), reflecting more effortful reading. Moreover, the effect of total cross-language (L2) orthographic neighborhood density was significant for both gaze duration ( $\beta = -31.43$ ,  $SE = 10.86$ ,  $t = -2.90$ ,  $p = 0.006$ ) and total reading time ( $\beta = -43.88$ ,  $SE = 16.64$ ,  $t = -2.64$ ,  $p = 0.008$ ). Words with many vs. few cross-language neighbors were easier to process, evidenced by shorter gaze durations (276 vs. 306 ms) and total reading times (370 vs. 430 ms) (see text footnote 2). The interactions with age group and language group were non-significant for both eye movement measures.

#### **Summary of Model 3's Findings**

We observed facilitatory total cross-language (L2) orthographic neighborhood density effects across both reading stages that were age-invariant and language background-invariant.

#### **Model 4: Cross-Language (L1) Neighborhood Density Effects on L2 Reading**

Only the bilingual groups were included in this analysis. The fixed factors were the same as those in Model 2, except that total cross-language (L1) orthographic neighborhood density (continuous) was included instead. The word-related and participant-related control predictors were also the same as those in Model 2, with the addition of mean cross-language bigram frequency (continuous, log-transformed), total cross-language phonological neighborhood density (continuous), cross-language orthographic neighborhood density of higher-frequency neighbors (continuous), and cross-language orthographic neighborhood frequency of higher-frequency neighbors (continuous). The random effects were the same as those in previous models. Complete model outputs for this analysis can be found in **Supplementary Table A6**.

The effect of total cross-language (L1) orthographic neighborhood density was significant for gaze duration ( $\beta = -144.19$ ,  $SE = 65.34$ ,  $t = -2.21$ ,  $p = 0.032$ ). Words with many vs. few within-language neighbors were easier to process, evidenced by shorter gaze durations (290 vs. 425 ms) (see text footnote 2). The interaction with age group was non-significant for both eye movement measures.

#### **Summary of Model 4's Findings**

We observed facilitatory total cross-language (L1) orthographic neighborhood density effects during early stage reading that were age-invariant.

## **DISCUSSION**

While engaging in visual word recognition, individuals must access and retrieve correct lexical representations from their mental lexicon among an array of visually similar lexical candidates: orthographic neighbors. While this process is limited to the activation of within-language orthographic neighbors among monolinguals, it is more complex among bilinguals due to the simultaneous activation of both within-language and cross-language orthographic neighbors. Thus, bilingual visual word recognition is influenced by the non-selective activation of both target and non-target language lexical representations, even in unilingual contexts.

While much is known about orthographic neighborhood effects among monolingual young adults, far less is known among other age and language groups (e.g., children, bilinguals). With the overarching aim of developing a more comprehensive understanding of within-language and cross-language activation during naturalistic reading in diverse groups of people, the current study employed eye movement measures to examine how within-language and cross-language orthographic neighborhood density influence visual word recognition during L1 and L2 paragraph reading in groups of monolingual and bilingual children and young adults. We had four main findings: two pertained to L1 and L2 within-language effects and the other two pertained to L1 and L2 cross-language effects. Each finding is discussed in turn.

#### **Within-Language (L1) Neighborhood Density Effects on L1 Reading**

Our first main finding was that high within-language (L1) orthographic neighborhood densities facilitated late-stage L1 word processing across all participant groups. Words with many within-language orthographic neighbors received shorter total reading times than those with fewer neighbors. However, the magnitude of these effects was larger in children, particularly the bilingual ones. Consistent with our original hypothesis, this finding suggests that the activation of multiple visually similar word forms facilitates target word recognition, particularly under conditions of low lexical entrenchment (i.e., when words have lower baseline activation levels and/or weaker links between different types of word-related information, as is likely the case for children and bilinguals). Given their reduced age, children's lexical representations have not benefited from as much life-long language exposure as those of young adults (their language and cognitive skills are still developing). Similarly, given their divided L1/L2 exposure, bilinguals' lexical representations have not benefited from as much absolute exposure as those of monolinguals. As a result, both conditions may entail reduced lexical entrenchment, evidenced by reduced ease of word processing. Combined, these conditions may engender a "double whammy," as evidenced by bilingual children's differentially reduced ease of word processing (see **Figure 1**). Accordingly, readers may maximally capitalize on high orthographic neighborhood densities under such conditions to identify the orthographic patterns of target words and retrieve

their meaning from the mental lexicon, in an effort to offset their reduced lexical accessibility (see, for example, Laxon et al., 1988, 1994, 2002; Duñabeitia and Vidal-Abarca, 2008; for response-based studies reporting similar patterns among monolingual children). Indeed, there were no between-group differences in our participants' reading comprehension performance (the ultimate goal of reading), suggesting a compensatory reading strategy (see also Whitford and Titone, 2019, for a similar strategy among bilingual older adults during paragraph reading).

Although these findings do not support IA and BIA/BIA+’s original predictions (i.e., inhibitory orthographic neighborhood density effects, particularly for words with lower baseline activation levels and/or weaker links due to lateral inhibition from words with higher baseline activation levels and/or stronger links), they support their alternative interpretation: that the activation of multiple lexical candidates can boost the overall excitation of the mental lexicon which, in turn, can boost the familiarity and activation levels of target words.

With regard to the extant eye movement literature, our findings are inconsistent with the few monolingual studies that have reported inhibitory orthographic neighborhood effects during sentence reading in adults (e.g., Pollatsek et al., 1999, Experiment 2; Slattery, 2009). Rather, they are consistent with the few bilingual studies that have reported largely facilitatory within-language orthographic neighborhood effects during L1 novel (Dirix et al., 2017, Experiment 2) and paragraph (Whitford et al., 2016; Whitford and Titone, 2017, 2019) reading in adults. Thus, similar to the monolingual and bilingual response-based literatures, which have reported mixed patterns of facilitatory, inhibitory, and null orthographic neighborhood effects, as a function of different word processing tasks, our findings suggest that such effects can also differ during naturalistic reading, as a function of different reading tasks and goals. For example, reading numerous short unrelated sentences, followed by simple yes/no comprehension questions on a percentage of trials, could contribute to inhibitory effects, whereas reading lengthy paragraphs of text that place greater demands on the visual, executive functioning, and linguistic systems, followed by open-ended comprehension questions, could contribute to facilitatory effects. As such, readers may capitalize more on high orthographic neighborhood densities during more effortful reading conditions to offset the greater processing demands, a strategy similar in principle to that proposed earlier for the bilingual children.

## Within-Language (L2) Neighborhood Density Effects on L2 Reading

Our second main finding was that within-language (L2) neighborhood density effects were equivocal. Although we observed numerically facilitatory effects during late-stage L2 word processing across both age groups—a pattern that would support an alternative interpretation of BIA/BIA+ that can accommodate facilitatory within-language orthographic neighborhood effects, as well as the extant bilingual eye movement literature (Dirix et al., 2017, Experiment 2; Whitford et al., 2016; Whitford and Titone, 2017, 2019)—if

real, the effects are likely weak and may require more tightly controlled stimuli to isolate them. Thus, future work in this area is needed.

## Cross-Language (L2) Neighborhood Density Effects on L1 Reading

Our third main finding was that high cross-language (L2) orthographic neighborhood densities facilitated both early stage and late-stage L1 word processing across all participant groups. Words with many cross-language orthographic neighbors received shorter gaze durations and total reading times than those with fewer neighbors. Although we predicted larger facilitatory effects in bilingual children vs. bilingual adults, no interactions with age group or language group reached significance. We would like to highlight here that despite self-identifying as functionally monolingual, most of our monolingual participants did, however, have some minimal L2 (French) proficiency; they completed basic French courses through the Ontario educational curriculum. Thus, it is possible that their L2 proficiency was sufficient enough to experience cross-language activation of visually similar L2 word forms. We note, however, that the magnitude of this cross-language activation was numerically smaller than that experienced by the bilingual participants. A closer look at the participant group means (based on a median splits) (see text footnote 2), revealed larger facilitative effects among bilingual children (56 ms), followed by bilingual adults (28 ms), monolingual children (22 ms), and monolingual adults (16 ms). Nonetheless, we cannot rule out the possibility that these effects were driven by some other, uncontrolled factor.

On the whole, these findings support an alternative interpretation of BIA/BIA+ that can accommodate facilitatory cross-language orthographic neighborhood effects and are largely consistent with previous eye movement studies of orthographic neighborhood effects in bilingual adults (Dirix et al., 2017, Experiment 2; Whitford et al., 2016; Whitford and Titone, 2017, 2019).

## Cross-Language (L1) Neighborhood Density Effects on L2 Reading

Our fourth main finding was that high cross-language (L1) orthographic neighborhood densities facilitated early stage L2 word processing across both bilingual age groups. Again, words with many cross-language orthographic neighbors received shorter gaze durations than those with fewer neighbors. However, late-stage L2 word processing (i.e., total reading time) was not affected, suggesting that facilitation (potentially due to top-down semantic-to-lexical excitatory feedback) occurred sufficiently so during first pass reading of the target words. Although we also predicted larger facilitatory effects in children vs. adults, the magnitude was age-invariant. This is likely because the two bilingual age groups were matched (all  $p$ -values > 0.05) on objective measures of L2 reading proficiency: L2 WIAT-II Word Reading standard scores (88.55 vs. 81.18) and L2 WIAT-II Pseudoword Decoding standard scores (95.70 vs. 97.70). Moreover, despite having accrued less life-long L2 exposure, the children had significantly higher current L2 exposure levels than

the adults (39.70 vs. 12.73%;  $p < 0.001$ ). Thus, greater “in the moment” L2 experience levels or more “bilingual modes” could counteract any historically driven age-related differences in visual word recognition.

As with the other orthographic neighborhood density effects discussed earlier, these findings also support an alternative interpretation of BIA/BIA+ and are consistent with previous eye movement studies of orthographic neighborhood effects in bilingual adults (Dirix et al., 2017, Experiment 2; Whitford et al., 2016; Whitford and Titone, 2017, 2019).

## CONCLUSION

The current study represents the first systematic investigation of within-language and cross-language activation during reading by means of orthographic neighborhood effects in a number of relatively understudied groups, including monolingual children, bilingual children, and bilingual adults. This work makes important empirical and theoretical contributions to the field by demonstrating that visually similar word forms, both within and across languages, facilitates visual word recognition during reading conditions that resemble those encountered in everyday life. Future avenues of research should explore whether leading models of visual word recognition, such as IA and BIA/BIA+, which were originally developed for skilled adult readers processing isolated words, can simulate the observed pattern of findings.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The stimuli used in the study were copyrighted material: passages from the Reading Comprehension subtest of the Weschler Individual Achievement Test—Second Edition (WIAT-II; Weschler 2005). However, the data will be made available upon request by contacting the corresponding author.

## REFERENCES

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: activation or search? *J. Exp. Psychol.: Learn. Mem. Cogn.* 15, 802–814. doi: 10.1037/0278-7393.15.5.802
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: lexical similarity or orthographic redundancy? *J. Exp. Psychol.: Learn. Mem. Cogn.* 18, 234–254. doi: 10.1037/0278-7393.18.2.234
- Andrews, S. (1997). The effects of orthographic similarity on lexical retrieval: resolving neighborhood conflicts. *Psychonomic Bull. Rev.* 4, 439–461. doi: 10.3758/BF03214334
- Andrews, S., and Hersch, J. (2010). Lexical precision in skilled readers: individual differences in masked neighbor priming. *J. Exp. Psychol.* 139, 299–318. doi: 10.1037/a0018366
- Andrews, S., and Lo, S. (2012). Not all skilled readers have cracked the code: individual differences in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 152–163. doi: 10.1037/a0024953
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9780511801686

## ETHICS STATEMENT

The study involved human participants and was approved by Western University's Research Ethics Board (REB). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

VW designed the experiment, collected, processed, and analyzed the data, and wrote the manuscript. MFJ provided direct funding for the research and feedback during various stages of the research process, including critically evaluating the manuscript. Both authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Operating Grant awarded to MFJ (for participant payments), an NSERC Research Tools and Instruments Grant awarded to the Department of Psychology at Western University (for the eye-tracking equipment that was used), a Fonds de recherche du Québec—Nature et technologies (FRQNT) postdoctoral scholarship awarded to VW (to conduct the research), a New Brunswick Innovation Foundation (NBIF) start-up grant awarded to VW (to cover the open access publishing fees), and the Canada Research Chairs Program awarded to VW (to write the manuscript).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.674007/full#supplementary-material>

- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014
- Bates, D. M. (2007). *Linear Mixed Model Implementation in lme4*. Unpublished Manuscript. Madison, WI: University of Wisconsin.
- Bates, D. M., Kliegl, R., Vasishth, S., and Baayen, H. R. (2015). Parsimonious mixed models. *arXiv [preprint]*.
- Beauvillain, C. (1992). “Orthographic and lexical constraints in bilingual word recognition,” in *Cognitive Processing in Bilinguals*, ed. R. J. Harris (Amsterdam: Elsevier Science Publishers B.V.), 221–235. doi: 10.1016/S0166-4115(08)61497-1
- Bosma, E., and Nota, N. (2020). Cognate facilitation in Frisian-Dutch bilingual children's sentence reading: an eye-tracking study. *J. Exp. Child Psychol.* 189:104699. doi: 10.1016/j.jecp.2019.104699
- Brown, L., Sherbenou, R. J., and Johnsen, S. K. (1997). *Test of Nonverbal Intelligence - (TONI-III): Examiner's Manual*, 3rd Edn. Austin, TX: Pro-Ed.
- Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977

- Carreiras, M., Perea, M., and Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *J. Exp. Psychol.: Learn. Mem. Cogn.* 23, 857–871. doi: 10.1037/0278-7393.23.4.857
- Castles, A., Davis, C., and Forster, K. I. (2003). “Word recognition development in children: insights from masked-priming,” in *Masked Priming: State of the Art*, eds S. Kinoshita and S. Lupker (Hove: Psychology Press), 345–360.
- Coltheart, M., Davelaar, E., Jonasson, J. T., and Besner, D. (1977). “Access to the internal lexicon,” in *Attention and Performance*, ed. S. Dornic (Hilldale, NJ: Erlbaum), 535–555.
- Commissaire, E., Audusseau, J., and Casalis, S. (2019). Disentangling cross-language orthographic neighborhood from markedness effects in L2 visual word recognition. *Psychonomic Bull. Rev.* 26, 353–359. doi: 10.3758/s13423-018-1518-6
- Davis, C. J., and Lupker, S. J. (2006). Masked inhibitory priming in English: evidence for lexical inhibition. *J. Exp. Psychol.: Hum. Percept. Perform.* 32, 668–687. doi: 10.1037/0096-1523.32.3.668
- Davis, C. J., Perea, M., and Acha, J. (2009). Re(de)fining the orthographic neighborhood: the role of addition and deletion neighbors in lexical decision and reading. *J. Exp. Psychol.: Hum. Percept. Perform.* 35, 1550–1570. doi: 10.1037/a0014253
- Davis, C. J., and Taft, M. (2005). More words in the neighborhood: interference in lexical decision due to deletion neighbors. *Psychonomic Bull. Rev.* 12, 904–910. doi: 10.3758/BF03196784
- de Groot, A. M. B., Borgwaldt, S., Bos, M., and van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: language effects and task effects. *J. Mem. Lang.* 47, 91–124. doi: 10.1006/jmla.2001.2840
- Dijkstra, T., and van Heuven, W. J. B. (1998). “The BIA model and bilingual word recognition,” in *Localist Connectionist Approaches to Human Cognition*, eds J. Grainger and A. Jacobs (New York, NY: Mahwah), 189–225.
- Dijkstra, T., and van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingualism: Lang. Cogn.* 5, 175–197. doi: 10.1017/S1366728902003012
- Dirix, N., Cop, U., Drieghe, D., and Duyck, W. (2017). Cross-lingual neighborhood effects in generalized lexical decision and natural reading. *J. Exp. Psychol.: Learn. Mem. Cogn.* 43, 887–915. doi: 10.1037/xlm0000352
- Duñabeitia, J. A., Borragán, M., De Bruin, A., and Casaponsa, A. (2020). Changes in the sensitivity to language-specific orthographic patterns with age. *Front. Psychol.* 11:1691. doi: 10.3389/fpsyg.2020.01691
- Duñabeitia, J. A., Ivaz, L., and Casaponsa, A. (2016). Developmental changes associated with cross-language similarity in bilingual children. *J. Cogn. Psychol.* 28, 16–31. doi: 10.1080/20445911.2015.1086773
- Duñabeitia, J. A., and Vidal-Abarca, E. (2008). Children like dense neighborhoods: orthographic neighborhood density effects in novel readers. *Spanish J. Psychol.* 11, 26–35. doi: 10.1017/S113874160000408X
- Duyck, W., Desmet, T., Verbeke, L. P., and Brysbaert, M. (2004). WordGen: a tool for word selection and nonword generation in Dutch, English, German, and French. *Behav. Res. Methods Instr. Comput.* 36, 488–499. doi: 10.3758/BF03195595
- Forster, K. I., and Shen, D. (1996). No enemies in the neighborhood: absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *J. Exp. Psychol.: Learn. Mem. Cogn.* 22, 696–713. doi: 10.1037/0278-7393.22.3.696
- Frances, C., Navarra-Barindelli, E., and Martin, C. D. (2021). Inhibitory and facilitatory effects of phonological and orthographic similarity on L2 word recognition across modalities in bilinguals. *Sci. Rep.* 11:12812. doi: 10.1038/s41598-021-92259-z
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., and Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *J. Exp. Psychol.: Gen.* 140, 186–209. doi: 10.1037/a0022256
- Grainger, J. (1992). “Orthographic neighborhoods and visual word recognition,” in *Orthography, Phonology, Morphology, and Meaning*, eds R. Frost and L. Katz (Elsevier: Amsterdam), 131–146. doi: 10.1016/S0166-4115(08)62792-2
- Grainger, J., and Dijkstra, A. (1992). “On the representation and use of language information in bilinguals,” in *Cognitive Processing in Bilinguals*, ed. R. J. Harris (Amsterdam: Elsevier Science Publishers B.V.), 207–220. doi: 10.1016/S0166-4115(08)61496-X
- Grainger, J., and Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychol. Rev.* 103, 518–565. doi: 10.1037/0033-295X.103.3.518
- Grainger, J., and Ziegler, J. (2011). A dual-route approach to orthographic processing. *Front. Psychol.* 2:54. doi: 10.3389/fpsyg.2011.00054
- Gregg, J., and Inhoff, A. W. (2016). Misperception of orthographic neighbors during silent and oral reading. *J. Exp. Psychol.: Hum. Percept. Perform.* 42, 799–820. doi: 10.1037/xhp0000193
- Grossi, G., Savill, N., Thomas, E., and Thierry, G. (2012). Electrophysiological cross-language neighborhood density effects in late and early English-Welsh bilinguals. *Front. Psychol.* 3:408. doi: 10.3389/fpsyg.2012.00408
- Holcomb, P. J., Grainger, J., and O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *J. Cogn. Neurosci.* 14, 938–950. doi: 10.1162/089892902760191153
- Hollingshead, A. B. (1975). *Four-Factor Index of Social Status*. Unpublished manuscript. New Haven, CT: Yale University.
- Kroll, J. F., Gullifer, J. W., and Zirnstein, M. (2016). “Literacy in adulthood: reading in two languages,” in *Lifespan Perspectives on Bilingualism*, eds S. Montanari and E. Nicoladis (Washington, DC: American Psychological Association).
- Kuperman, V., Drieghe, D., Keuleers, E., and Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Q. J. Exp. Psychol.* 66, 563–580. doi: 10.1080/17470218.2012.658820
- Lauro, J., and Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: a meta-analytic review. *J. Mem. Lang.* 92, 217–233. doi: 10.1016/j.jml.2016.06.010
- Laxon, V., Coltheart, V., and Keating, C. (1988). Children find friendly words friendly too: words with many orthographic neighbours are easier to read and spell. *Br. J. Educ. Psychol.* 58, 103–119. doi: 10.1111/j.2044-8279.1988.tb00882.x
- Laxon, V., Gallagher, A., and Masterson, J. (2002). Word length and orthographic neighbourhood size effects in children's reading. *Br. J. Psychol.* 93, 269–287. doi: 10.1348/000712602162580
- Laxon, V., Masterson, J., and Moran, R. (1994). Are children's representations of words distributed? Effects of orthographic neighbourhood size, consistency, and regularity of naming. *Lang. Cogn. Processes* 9, 1–27. doi: 10.1080/01690969408402108
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., and Zwisterlood, P. (2008). Native language influences on word recognition in a second language: a megastudy. *J. Exp. Psychol.: Learn. Mem. Cogn.* 34, 12–31. doi: 10.1037/0278-7393.34.1.12
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230. doi: 10.1371/journal.pone.0043230
- Marian, V., Blumenfeld, K. H., and Kaushanskaya, M. (2007). The Language Proficiency and Experience Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marinus, E., and de Jong, P. F. (2010). Size does not matter, frequency does: sensitivity to orthographic neighbors in normal and dyslexic readers. *J. Exp. Child Psychol.* 106, 129–144. doi: 10.1016/j.jecp.2010.01.004
- Massol, S., Grainger, J., Dufau, S., and Holcomb, P. (2010). Masked priming from orthographic neighbors: an ERP investigation. *J. Exp. Psychol.: Hum. Percept. Perform.* 36, 162–174. doi: 10.1037/a0017614
- Mathey, S. (2001). Influence of orthographic neighborhood in word recognition tasks. *Can. J. Exp. Psychol./Revue Canadienne Psychol. Exp.* 55, 1–23.
- McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychol. Rev.* 88, 375–407. doi: 10.1037/0033-295X.88.5.375
- Meade, G., Midgley, K. J., Dijkstra, T., and Holcomb, P. J. (2018). Cross-language neighborhood effects in learners indicative of an integrated lexicon. *J. Cogn. Neurosci.* 30, 70–85. doi: 10.1162/jocn\_a\_01184
- Midgley, K. J., Holcomb, P. J., Van Heuven, W. J. B., and Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Res.* 1246, 123–135. doi: 10.1016/j.brainres.2008.09.078

- Miellet, S., Sparrow, L., and Sereno, S. C. (2007). Word frequency and predictability effects in reading French: an evaluation of the E-Z Reader model. *Psychonomic Bull. Rev.* 14, 762–769. doi: 10.3758/BF03196834
- Mulder, K., van Heuven, W. J., and Dijkstra, T. (2018). Revisiting the neighborhood: how L2 proficiency and neighborhood manipulation affect bilingual processing. *Front. Psychol.* 9:1860. doi: 10.3389/fpsyg.2018.01860
- Nakayama, M., Sears, C. R., and Lupker, S. J. (2008). Masked priming with orthographic neighbors: a test of the lexical competition assumption. *J. Exp. Psychol.: Hum. Percept. Perform.* 34, 1236–1260. doi: 10.1037/0096-1523.34.5.1236
- New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Use base de données lexicales du français contemporain sur internet: lexique 3. *L'Année Psychol.* 101, 447–462. doi: 10.3406/psy.2001.1341
- Perea, M., and Estévez, A. (2008). Transposed-letter similarity effects in naming pseudowords: evidence from children and adults. *Eur. J. Cogn. Psychol.* 20, 33–46. doi: 10.1080/09541440701306941
- Perea, M., and Fraga, I. (2006). Transposed-letter and laterality effects in lexical decision. *Brain Lang.* 97, 102–109. doi: 10.1016/j.bandl.2005.08.004
- Perea, M., and Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *J. Exp. Psychol.: Hum. Percept. Perform.* 24, 767–779. doi: 10.1037/0096-1523.24.3.767
- Perea, M., and Rosa, E. (2000). The effects of orthographic neighborhood in reading and laboratory word identification tasks: a review. *Psicológica* 21, 327–340.
- Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730
- Pollatsek, A., Perea, M., and Binder, K. (1999). The effects of “neighborhood size” in reading and lexical decision. *J. Exp. Psychol.: Hum. Percept. Perform.* 25, 1142–1158. doi: 10.1037/0096-1523.25.4.1142
- Pollatsek, A., Reichle, E. D., and Rayner, K. (2006). Tests of the E-Z Reader model: exploring the interface between cognition and eye movement control. *Cogn. Psychol.* 52, 1–56. doi: 10.1016/j.cogpsych.2005.06.001
- R Development Core Team (2021). *R: A Language and Environment for Statistical Computing*. (Version 3.6.1). R Foundation for Statistical Computing, Vienna, Austria. Vienna: R Development Core Team.
- Radach, R., Huestegge, L., and Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychol. Res.* 72, 675–688. doi: 10.1007/s00426-008-0173-3
- Rau, A. K., Moll, K., Moeller, K., Huber, S., Snowling, M. J., and Landerl, K. (2016). Same same, but different: word and sentence reading in German and English. *Sci. Stud. Read.* 20, 203–219. doi: 10.1080/10888438.2015.1136913
- Rau, A. K., Moll, K., Snowling, M. J., and Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *J. Exp. Child Psychol.* 130, 92–105. doi: 10.1016/j.jecp.2014.09.012
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461
- Schröter, P., and Schroeder, S. (2016). Orthographic processing in balanced bilingual children: cross-language evidence from cognates and false friends. *J. Exp. Child Psychol.* 141, 239–246. doi: 10.1016/j.jecp.2015.09.005
- Schröter, P., and Schroeder, S. (2018). Exploring early language detection in balanced bilingual children: the impact of language-specificity on cross-linguistic nonword recognition. *Int. J. Bilingualism* 22, 305–315. doi: 10.1177/1367006916672751
- Sears, C. R., Campbell, C. R., and Lupker, S. J. (2006). Is there a neighbor frequency effect in English? Evidence from reading and lexical decision. *J. Exp. Psychol.: Hum. Percept. Perform.* 32, 1040–1062. doi: 10.1037/0096-1523.32.4.1040
- Sears, C. R., Hino, Y., and Lupker, S. J. (1995). Neighborhood frequency and neighborhood size effects in visual word recognition. *J. Exp. Psychol.: Hum. Percept. Perform.* 21, 876–900.
- Sebastián-Gallés, N., and Parraño, A. (1995). The development of analogical reading in Spanish. *Read. Writ.* 7, 23–38. doi: 10.1007/BF01026946
- Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: evidence from eye movements. *J. Exp. Psychol.: Hum. Percept. Perform.* 35, 1969–1975. doi: 10.1037/a0016894
- Snodgrass, J. G., and Mintzer, M. (1993). Neighborhood effects in visual word recognition: facilitatory or inhibitory? *Mem. Cogn.* 21, 247–266. doi: 10.3758/BF03202737
- Tamura, N., Castles, A., and Nation, K. (2017). Orthographic learning, fast and slow: lexical competition effects reveal the time course of word learning in developing readers. *Cognition* 163, 93–102. doi: 10.1016/j.cognition.2017.03.002
- Titone, D., Whitford, V., Lijewska, A., and Itzhak, I. (2016). “Bilingualism, executive control, and eye movement measures of reading: a selective review and reanalysis of bilingual vs. multilingual reading data,” in *Cognitive Control and Consequences in the Multilingual Mind*, ed. J. Schwieter (Amsterdam: John Benjamins), 11–46. doi: 10.1075/bpa.2.02tit
- Van Assche, E., Duyck, W., and Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Front. Psychol.* 3:174. doi: 10.3389/fpsyg.2012.00174
- van Hell, J. G. (2020). “Lexical processing in child and adult classroom second language learners: uniqueness and similarities, and implications for cognitive models,” in *Psychology of Learning and Motivation*, Vol. 72, eds K. D. Federmeier, and H.-W. Huang (Cambridge, MA: Academic Press), 207–234.
- Van Heuven, W. J. B., Dijkstra, T., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Mem. Lang.* 39, 458–483. doi: 10.1006/jmla.1998.2584
- Wechsler, D. (2005). *Wechsler Individual Achievement Test – (WIAT-II)*, 2nd Edn. London: The Psychological Corporation.
- Whitford, V., and Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children’s first-language and second-language reading? A focus on word frequency effects. *J. Exp. Child Psychol.* 173, 318–337. doi: 10.1016/j.jecp.2018.03.014
- Whitford, V., Pivneva, I., and Titone, D. (2016). “Eye movement methods to investigate bilingual reading,” in *Methods in Bilingual Reading Comprehension Research*, eds R. R. Heredia, J. Altarriba, and A. B. Cieslicka (New York, NY: Springer), 183–211. doi: 10.1007/978-1-4939-2993-1\_8
- Whitford, V., and Titone, D. (2012). Second language experience modulates first- and second-language word frequency effects: evidence from eye movement measures of natural paragraph reading. *Psychonomic Bull. Rev.* 19, 73–80. doi: 10.3758/s13423-011-0179-5
- Whitford, V., and Titone, D. (2014). The effects of reading comprehension and launch site on frequency-predictability interactions during paragraph reading. *Q. J. Exp. Psychol.* 67, 1151–1165. doi: 10.1080/17470218.2013.848216
- Whitford, V., and Titone, D. (2017). The effects of word frequency and word predictability during first- and second-language paragraph reading in bilingual older and younger adults. *Psychol. Aging* 32, 158–177. doi: 10.1037/pag0000151
- Whitford, V., and Titone, D. (2019). Lexical entrenchment and cross-language activation: two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Lang. Cogn.* 22, 58–77. doi: 10.1017/S1366728917000554
- Ziegler, J. C., and Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Dev. Sci.* 9, 429–453.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., and Schulte-Körne, G. (2003). Developmental dyslexia in different languages: language-specific or universal? *J. Exp. Child Psychol.* 86, 169–193. doi: 10.1016/S0022-0965(03)00139-5

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Whitford and Joanisse. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Child Heritage Language Development: An Interplay Between Cross-Linguistic Influence and Language-External Factors

Natalia Meir<sup>1\*</sup> and Bibi Janssen<sup>2</sup>

<sup>1</sup> Department of English Literature and Linguistics, Bar-Ilan University, Ramat Gan, Israel, <sup>2</sup> Independent Researcher, Amsterdam, Netherlands

## OPEN ACCESS

### Edited by:

Zofia Wodniecka,  
Jagiellonian University, Poland

### Reviewed by:

Joanna Kolak,  
University of Warsaw, Poland  
Jennifer Byrnes Austin,  
Rutgers University, Newark,  
United States

### \*Correspondence:

Natalia Meir  
natalia.meir@biu.ac.il

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 January 2021

**Accepted:** 27 September 2021

**Published:** 11 November 2021

### Citation:

Meir N and Janssen B (2021) Child Heritage Language Development: An Interplay Between Cross-Linguistic Influence and Language-External Factors. *Front. Psychol.* 12:651730. doi: 10.3389/fpsyg.2021.651730

The current study investigated the mechanisms of heritage language (HL) development with a focus on case morphology. First, the effects of cross-linguistic influence (i.e., the influence of the properties of the societal language (SL) on the acquisition of the HL) was assessed by performing bilingual vs. monolingual, and between-bilingual group comparisons (Russian–Dutch vs. Russian–Hebrew bilinguals). Russian, Hebrew, and Dutch show differences in the marking of the accusative (ACC) and genitive (GEN) cases, and these differences were used as a basis for the evaluation of cross-linguistic influences. Second, the study evaluated the contribution of language-external factors such as chronological age, age of onset of bilingualism (AoO), languages spoken by the parent to the child (only HL, only SL, both HL and SL), and family language type (both parents are HL speakers, mixed families). Finally, we assessed how language-external factors might potentially mitigate the effects of cross-linguistic influences in bilinguals. Russian–Dutch bilinguals from the Netherlands ( $n = 39$ ,  $M_{AGE} = 5.1$ ,  $SD = 0.8$ ), Russian–Hebrew bilinguals from Israel ( $n = 36$ ,  $M_{AGE} = 4.9$ ,  $SD = 0.9$ ) and monolingual Russian-speaking children ( $n = 41$ ,  $M_{AGE} = 4.8$ ,  $SD = 0.8$ ), along with adult controls residing in the Russian Federation, participated in the study. The case production of ACC and GEN cases was evaluated using elicitation tasks. For the bilinguals, the background data on individual language-external factors were elicited from the participants. The results show that case morphology is challenging under HL acquisition—case acquisition in the HL is impeded under the influence of the properties of the SL. This is evident in the lower performance of both bilingual groups, compared with the monolingual controls who showed ceiling performance in the production of target inflection in the ACC and GEN contexts. More specifically, the acquisition of morphology is hindered when there are differences in the mapping of functional features (such as with Russian–Hebrew bilinguals) and/or the absence of this feature marking (such as with Russian–Dutch bilinguals). But the findings also point to the involvement of language-external factors as important mitigators of potential negative effects of cross-linguistic influence. In summary, HL development is an intricate interplay between cross-linguistic influence and language-external factors.

**Keywords:** heritage language development, child bilingualism, cross-linguistic influence, input, case morphology, Russian, Hebrew, Dutch

## INTRODUCTION

The term “heritage language” (HL)—also called “minority language,” “community language,” “home language,” “family language,” “mother tongue,” and “L1”—refers to a language that is spoken at home, but not by the majority of society which uses societal language (SL) (Rothman, 2009; Benmamoun et al., 2013; Montrul, 2016; Polinsky, 2018; Ortega, 2020). Numerous studies have investigated the end-state grammars of adult HL speakers (for an overview see Montrul, 2016; Polinsky, 2018; Polinsky and Scontras, 2019, 2020). The current study aims to contribute to the understanding of the mechanisms affecting child HL development. We used the term “child HL speakers,” but it should be noted that in previous research, the term “L1 development” in simultaneous or early sequential bilinguals has been used instead [for a detailed discussion on the terminology in HL and child bilingualism research see Kupisch and Rothman (Kupisch and Rothman, 2018)]. Today, more and more studies use the term “child HL speakers” (e.g., Meir and Armon-Lotem, 2015; Cuza and Pérez-Tattam, 2016; Daskalaki et al., 2019, 2020; Chondrogianni and Schwartz, 2020; Goebel-Mahrle and Shin, 2020; Rodina et al., 2020; Serratrice, 2020; Armon-Lotem et al., 2021; Otwinowska et al., 2021). Heritage language speakers acquire their HL from birth *via* naturalistic input, but as adults, they show divergence from the baseline (the language spoken in the country of origin, or the language spoken by the first generation of immigrants who are dominant in this language). Among the possible mechanisms triggering divergences in HL grammars, researchers have proposed cross-linguistic influences from the SL and/or language-external factors (such as the age of onset of bilingualism (AoO), quantity and quality of exposure to HL and SL, family type (both parents are HL speakers vs. a mixed HL-SL family), HL community size, and the number of HL speakers in the child’s environment). For a detailed review of the potential mechanisms affecting HL development see Montrul (2016), Polinsky (2018), and Polinsky and Scontras (2019, 2020).

There is no agreement on the underlying mechanisms affecting morphosyntactic development in child HL speakers. Some studies reported that child HL speakers pair up with monolingual peers on morphosyntax. Alternatively, studies showed that child HL speakers diverge from their monolingual peers in this domain. Some researchers attribute these divergences to cross-linguistic influences (for a detailed overview on cross-linguistic influence in child bilingualism see Serratrice, 2013; Van Dijk et al., 2021), while some studies found that cross-linguistic influences did not affect the HL acquisition of morphosyntax, suggesting that language-external factors shape child HL development (e.g., Daskalaki et al., 2019, 2020; Rodina et al., 2020). Cross-linguistic influences and language-external factors might not be mutually exclusive and shape HL development together (Daskalaki et al., 2019, 2020; Van Dijk et al., 2021). In the current study, we evaluate the involvement of both factors.

To contribute to this ongoing debate, two groups of bilinguals, who acquired Russian as their HL in contact with two typologically different SLs (Dutch and Hebrew), were compared with their monolingual Russian-speaking peers regarding the

production of case inflectional morphology. We compared the bilinguals with their monolingual peers, and we compared the two bilingual groups which acquired the same HL with typologically different SLs. The choice of case inflectional morphology was motivated by the fact that Russian is a language with a rich case morphology that realizes case features morphologically. Dutch does not mark cases on nouns, while Hebrew marks the accusative (ACC) cases with the dedicated particle “*et*” in front of definite noun phrases. Neither Dutch nor Hebrew uses special morphology to mark genitive (GEN) cases in negative constructions. These properties of SL-Dutch and SL-Hebrew might potentially impede the acquisition and maintenance of case morphology in HL-Russian of bilingual children. Thus, the comparison of two different bilingual groups, whose SLs differ in marking cases, is expected to shed light on whether SL properties affect the HL development of a child, and how language-external factors shape HL development and mitigate the potential negative effects of the SL on the HL.

In this introductory section, we first review studies assessing cross-linguistic influence in HL development. Second, we discuss the effects of language-external factors on HL development. Third, we provide a brief overview of the case morphology in the Russian Language (the HL of the bilingual children in the current study), and then we discuss the typological differences between Russian, Dutch, and Hebrew focusing on case marking in the three languages. We conclude the introductory section with the research questions and hypotheses of the current study.

## Factors Affecting Child Heritage Language (HL) Acquisition

### Cross-Linguistic Influence in Child HL Acquisition

Cross-linguistic influence has been proposed to account for the success or failure of bilinguals in their acquisition of certain linguistic properties in one language in the presence of the other language. A facilitative effect is observed when the properties of the two languages have converging configurations. A negative cross-linguistic influence (also called cross-linguistic transfer/interference), defined as “deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language” (Weinreich, 1968), has been robustly demonstrated in the acquisition of L2/SL in simultaneous and sequential bilingual children (e.g., Zdorenko and Paradis, 2008; Blom et al., 2012). A recent meta-analysis which evaluated cross-linguistic influence and its predictors in 750 simultaneous and early sequential bilingual children (aged 4;0–10;0) in 17 unique language combinations across 26 experimental studies confirmed the presence of cross-linguistic influences in bilingual morphosyntactic acquisition (see Van Dijk et al., 2021).

Several hypotheses have been suggested to account for the cross-linguistic influences in various bilingual child and adult populations (Blom et al., 2017). Some propose that cross-linguistic influence occurs when there is a partial structural overlap between the two languages, while cross-linguistic influence is not predicted when the language structures are

either completely distinct or completely overlap (Hulk and Müller, 2000). The *feature (re)-assembly hypothesis* (Lardiere, 2009) proposes that not only the absence/presence of a feature in the two languages determines learnability but also the differences in the mapping and bundling of features in the L1 and the L2 shape the trajectory of acquisition. A facilitative effect is expected if the properties are similarly mapped in the L1 and L2, while an impeding effect is expected when the properties in the two languages have different configurations. The predictions of some cross-linguistic influence hypotheses, originally proposed for L2 acquisition, have been extended to adult and child HL acquisition as well. For example, the *feature (re)-assembly hypothesis* (Lardiere, 2009) has been extended to child HL acquisition (see Cuza and Pérez-Tattam, 2016; Meir et al., 2017), showing that feature (re)-assembly in the HL is affected by the properties of the SL.

Individual studies evaluating the effects of cross-linguistic influence on child HL acquisition provide inconclusive evidence. There is accumulating evidence that SL properties affect HL development (e.g., Guiberson et al., 2006; Gathercole and Thomas, 2009; Meir et al., 2017). Most of the previous studies evaluating cross-linguistic influence in HL development have employed a monolingual vs. bilingual comparison paradigm. For example, Montrul and Sánchez-Walker (2013) showed evidence for the influence of SL-English on HL-Spanish in child heritage speakers aged 6;0–17;0 for the Spanish differential object marker *a*. Child heritage speakers of Spanish were reported to omit the differential object marker significantly more frequently than their monolingual peers. This has been linked to the influence of SL-English which does not have a differential object marker. Similarly, Cuza and Pérez-Tattam (2016) compared two groups of Spanish-speaking children (aged 5;0–10;8): HL-Spanish and SL-English speakers, and monolingual controls. The authors reported significant differences between the two groups. Furthermore, the authors attributed the decreased accuracy on gender agreement and gender assignment in HL-Spanish to the properties of SL-English which do not mark grammatical gender. In the same vein, Meir et al. (2017) compared HL-Russian speakers with their monolingual controls and reported an asymmetry in the performance on ACC case accuracy and subject-verb agreement accuracy. The HL speakers paired up with their monolingual peers on subject-verb agreement, but they were less accurate in terms of the ACC case morphology. This asymmetry in their performance has been linked to the properties of SL-Hebrew; Hebrew marks gender, number, and person features on verbal inflections similarly to Russian, while the marking of case morphology is different in the two languages. Again, the lower accuracy on the ACC case of HL-Russian speakers was linked to the SL properties, suggesting cross-linguistic influence. It is also important to note that bilingual children might show an asymmetry between the comprehension and production of the same grammatical phenomenon. For example, Kim et al. (2018) showed that HL-Korean bilingual children aged 8;0–10;0 were less accurate in their comprehension of object-subject-verb (OSV) sentences as compared with their monolingual peers. The gap in comprehension might be attributed to the lack of

ACC marking in English and the word order properties of English. But on the production tasks, the same children correctly produced the ACC marking. A meta-analysis by Van Dijk et al. (2021) compared the magnitude of cross-linguistic influence in 26 studies between bilinguals and their monolingual peers on a wide range of morphosyntactic structures and concluded that “cross-linguistic influence is part and parcel of being bilingual and can manifest itself in various linguistic contexts.” The authors furthermore concluded that the magnitude of cross-linguistic influence is stronger from the SL to the HL, which is crucial for the predictions of the current study, as our study focuses on the influence of two typologically different SLs on HL-Russian.

Another paradigm that is used to evaluate cross-linguistic influence is the comparison of several bilingual groups, rather than the bilingual vs. monolingual comparison (e.g., Sorace et al., 2009; Schwartz et al., 2015; Janssen and Meir, 2019; Kaltsa et al., 2020; Rodina et al., 2020). Such studies were not included in the meta-analysis by Van Dijk et al. (2021) due to their scarcity. Studies employing multiple bilingual group comparisons bring inconclusive results. On the one hand, there are studies confirming the influence of SL on HL. For example, Schwartz et al. (2015) compared the noun–adjective gender agreement in HL-Russian with various SLs (English, Finnish, German, and Hebrew). The authors reported the cross-linguistic influence of the SL properties on HL gender acquisition/maintenance. The HL-Russian speakers whose SL marks grammatical gender (German, Hebrew) outperformed the bilinguals whose SL does not mark grammatical gender (English, Finnish) on gender agreement in HL-Russian. On the other hand, a study by Rodina et al. (2020) also compared the noun–adjective gender agreement in HL-Russian speakers with varying SLs (German, Hebrew, Norwegian, Latvian, and English) and found that cross-linguistic influence did not affect HL acquisition. Similar to the study by Schwartz et al. (2015), the study by Rodina et al. (2020) included languages that had similar configurations to Russian concerning gender assignment/markings and languages which did not mark grammatical gender. Russian has a three-way gender system, as do German and Norwegian; Hebrew and Latvian have two-gender systems, while English does not mark grammatical gender at all. Yet the study by Rodina et al. (2020) found no evidence for the influence of SL on HL development. Rather, language-external factors were reported to shape HL grammatical gender acquisition.

To sum up, based on monolingual vs. bilingual comparisons, the presence of cross-linguistic influence has been shown for both SL and HL acquisition (see Van Dijk et al., 2021). More specifically, the lower accuracy on morphosyntax in the HL may be traced back to the influence of the SL, and the magnitude of cross-linguistic influence is higher from the SL to the HL, as compared with the cross-linguistic influence in the opposite direction. However, some studies show no effect of cross-linguistic influence, especially based on between-bilingual-group comparisons, but such studies are less frequent. The latter studies show that language-external variables rather than cross-linguistic

influence shape HL development. In the next subsection, we focus on the role of language-external variables in HL development.

### Language-External Factors in Child HL Acquisition

Language-external factors shape monolingual as well as bilingual language acquisition in both languages (see Unsworth, 2013, 2014, 2016, 2019; De Houwer et al., 2018; Armon-Lotem and Meir, 2019). The effects of various language-external factors have been evaluated in HL development: chronological age, longer period of uninterrupted HL acquisition (as indexed by AoO), greater use of HL at home, the size of the HL community, and current and cumulative exposure to HL predict the success of HL development (e.g., Daskalaki et al., 2019, 2020; Rodina et al., 2020; Armon-Lotem et al., 2021; Otwinowska et al., 2021), yet the specific details of the findings vary.

We begin our discussion with chronological age. Although the effect of chronological age is robust in monolingual typical language development, it is not always observed in bilingual HL development (Lein et al., 2017; Gagarina and Klassert, 2018). On one hand, studies have shown that HL skills improve with age, but on the other hand, it has also been demonstrated that HL competence shows fossilization and even declines with age. For example, Gagarina and Klassert (2018) showed the effects of age for the acquisition of HL-Russian morphosyntax in Russian-German bilinguals. Additionally, Cuza et al. (2021) showed that while child HL-Spanish speakers diverge on copular production, this divergence is diminished in adult HL speakers. However, a meta-analysis investigating the effects of cross-linguistic influence in child morphosyntactic acquisition found that this influence did not significantly change over time (see Van Dijk et al., 2021), suggesting that the effect of cross-linguistic influence is present in older bilingual children as well.

Furthermore, chronological age in bilinguals might be related to input/exposure characteristics and language dominance characteristics: some HL speakers continue to receive input in their HL, some HL speakers receive more input/exposure in the SL. Indeed, HL input/exposure quantity is an important predictor of the success/failure of HL acquisition (for an overview see Armon-Lotem and Meir, 2019). The input/exposure of HL diminishes as HL speakers get older and socialize using the SL (e.g., Unsworth, 2014; De Houwer et al., 2018), which might explain the fossilization and attrition/loss over time. For example, the HL input quantity at home is related to HL development: child HL-Greek speakers with <25% of HL use had very low accuracy rates on the morphosyntactic measures (Daskalaki et al., 2019). The AoO of SL is linked to HL morphosyntactic development: children with later SL AoOs, i.e., children with longer periods of uninterrupted HL acquisition, are shown to perform better on the morphosyntax compared with HL child speakers with earlier AoOs of SL (Meir et al., 2017; Armon-Lotem et al., 2021). Rodina et al. (2020) showed that several language-external factors such as family type (mixed family vs. HL-speaking parents), age, and current exposure to HL instruction are related to the HL development of grammatical gender in HL speakers with various SL properties. In contrast, Tsiniivits and Unsworth (2021) showed that having an older sibling is

related to the increased input of SL in the home setting, and hence higher SL abilities, yet having an older sibling was reported to have no effect on HL grammatical complexity. It should be noted that the input/exposure factors are related to language dominance/proficiency—children with more exposure/input to the language show higher dominance in that language (e.g., Treffers-Daller, 2019).

In addition to the diminished input quantity that HL speakers receive compared with their monolingual peers, the input quality is shown to be different from what monolinguals receive in the country where their language is dominant (e.g., Rothman, 2007; Montrul and Sánchez-Walker, 2013). HL speakers are not schooled in their HL; thus, their HL input is limited to home contexts (Montrul, 2016; Polinsky, 2018). For example, Daskalaki et al. (2020) showed that child HL-Greek speakers indeed diverge from monolingual Greek-speaking children on subject placement. The divergence from the monolingual controls was attributed not only to the decreased input that the HL-Greek speakers were receiving compared with monolingual controls but also to the divergences which were already visible in the speech of the parents of these children, showing that HL speakers are exposed to divergent inputs.

Thus, numerous language-external factors have been proposed to account for HL development in bilingual children. In the current study, we tested the contribution of cross-linguistic influences and language-external factors to shed light on the mechanism of HL development. Cross-linguistic influences and language-external factors might not be mutually exclusive (Daskalaki et al., 2020), and all contribute to HL development. Indeed, some scholars advocate that “input (quality/quantity) is not the only factor determining HL acquisition” (see Putnam and Sánchez, 2013, p. 487). Therefore, in the current study, we will consider both factors, language-external factors which index the quantity and quality of input, as well as cross-linguistic influences. Putnam and Sánchez (2013) proposed that the (re)-assembly of features in the HL (Lardiere, 2009) is related to the activation of a specific HL feature in production and comprehension. The HL feature activation might be related to cross-linguistic and language-external competitions which might result in its insufficient activation and gradual replacement by SL features. Before presenting the research questions of the study, we will briefly review the case morphology in Russian, Dutch, and Hebrew, which is the phenomenon that serves as a test case for the comparison between the monolingual and HL morphosyntactic development.

### Case Morphology in Russian, Dutch, and Hebrew

Russian is a language with a rich nominal inflectional morphology. Russian realizes case features morphologically: all Russian nouns, adjectives, quantifiers, demonstratives, and numerals must bear a dedicated case inflection (e.g., Bailyn, 2012). There are six main cases in Russian in singular and plural: nominative (NOM), GEN, ACC, dative (DAT), instrumental (INSTR), and prepositional (PREP). In the current study, we will

focus on two cases: the ACC case of direct objects and the GEN case in negative constructions.

The choice of ACC and GEN case forms for the current study was not accidental. While Russian has a rich nominal morphology, Dutch and Hebrew have no case inflections. In Hebrew, the use of the ACC marking is limited (only in front of definite nouns), while in Dutch there are no special markings in the investigated constructions. These properties of the three languages will serve as a testing point to shed light on the mechanisms of HL development.

**Table 1** lists the ACC and GEN case paradigms for singular nouns across the first and second declension classes for animate and inanimate nouns. In some nouns, the ACC and GEN case marking is visible, i.e., ACC/GEN are non-homophonous to NOM ones. For example, the noun *devočka* “girl” has three different forms for the NOM—ACC—GEN (e.g., *devočk-a* - *devočk-u* - *devočk-i*). However, Russian is characterized by extensive morphological case syncretism. For example, for some nouns, ACC and GEN forms are homophonous to their NOM forms: the noun *mjaso* “meat” has three forms that overlap in their phonological form. For the word *stol* “table,” the ACC and the NOM forms overlap, while the GEN form is different (e.g., *stol* - *stol* - *stol-a*).

Nominative case forms are the most frequent in the Russian National Corpus (30.5%), followed by ACC forms, which comprise 20% of the total nominal inflections and 23% for GEN forms for singular nouns (as per the frequency data provided by Slioussar and Samoilova, 2015). Thus, it is plausible to suggest that form frequency should not trigger differences in the acquisition trajectory in monolingual and bilingual children. To the best of our knowledge, there is no available frequency analysis for the case form distributions in monolingual and bilingual Russian child-directed speech. Yet, looking into the data on Polish, a typologically close language to Russian, in early childhood (around age 2), about 54% of the nouns in child-directed speech are in the NOM form, 19% in the ACC, and 12% in the GEN (Dąbrowska and Szczerbiński, 2006). It is plausible to suggest that these percentages observed for Polish might be

similar in Russian as well, yet future studies should investigate the distribution of different case forms in Russian monolingual and bilingual child-directed speech.

Having presented the Russian case system, we now turn to Dutch and Hebrew, both of which have sparse case morphology, but the two languages show fundamental differences with respect to case marking. Dutch does not mark cases on nouns, whereas Hebrew marks the ACC case by the particle *et* only before definite nouns (Berman, 1978) and uses possessive GEN inflectional markers. In negative constructions, there is no special marker on the noun in Hebrew. Below, we present a comparison between Russian, Dutch, and Hebrew for the ACC (see 1) and GEN (see 2) environments investigated in the current study.

(1) ACC case (Direct object)

|         |                                  |                             |                              |
|---------|----------------------------------|-----------------------------|------------------------------|
| Russian | <i>mal'čik</i><br><b>boy.NOM</b> | <i>vidit</i><br>sees        | <i>devočk-u</i><br>girl-ACC  |
| Dutch   |                                  |                             |                              |
| (INDEF) | <i>de jongen</i><br>DEF.boy      | <i>ziet</i><br>sees.SG.3P   | <i>een</i><br>INDEF          |
| (DEF)   | <i>de jongen</i><br>DEF.boy      | <i>ziet</i><br>sees.SG.3P   | <i>het</i><br>DEF            |
| Hebrew  |                                  |                             |                              |
| (INDEF) | <i>ha- yeled</i><br>DEF.boy      | <i>ro'e</i><br>sees.M.SG.3P | <i>yalda</i><br>girl         |
| (DEF)   | <i>ha- yeled</i><br>DEF.boy      | <i>ro'e</i><br>sees.M.SG.3P | <i>et</i><br>ACC             |
|         |                                  |                             | <i>ha- yalda</i><br>DEF-girl |

(2) GEN case (Negative constructions)

|         |                             |                      |                      |                             |
|---------|-----------------------------|----------------------|----------------------|-----------------------------|
| Russian | <i>a</i><br>and             | <i>teper'</i><br>now | <i>net</i><br>no     | <i>devočk-i</i><br>girl-GEN |
| Dutch   | <i>en</i><br>and            | <i>nu</i><br>now     | <i>niet</i><br>no    | <i>meisje</i><br>girl       |
| Hebrew  | <i>ve-axšav</i><br>and- now | <i>ein</i><br>no     | <i>yalda</i><br>girl |                             |

## RUSSIAN CASE ACQUISITION AMONG MONOLINGUAL AND BILINGUAL CHILDREN

Monolingual Russian-speaking children acquire the basics of the case system within a very short period (e.g., Gvozdev, 1961; Babyonyshev, 1993; Protassova and Voeikova, 2007; Cejtin, 2009). Based on previous case studies, it is known that the NOM—ACC oppositions occur at about 1;9 (Gvozdev, 1961). Babyonyshev (Babyonyshev, 1993) showed that monolingual Russian-speaking children have full mastery of NOM (597 out of 600) and ACC (27 out of 30) cases from the moment of appearance of the arguments that require them. Furthermore, Babyonyshev (1993) found no errors with the GEN of negation in the speech of monolingual Russian-speaking children. Babyonyshev (1993) argued that the use of NOM, ACC, and GEN (in negation constructions) is already operative in monolingual Russian-speaking 2-year-olds.

**TABLE 1 |** The Russian case inflections ([NOM] → [ACC] → [GEN]) across 2 declension classes for animate and inanimate nouns.

|                   |                  | English<br>Translation | NOM            | ACC            | GEN            |
|-------------------|------------------|------------------------|----------------|----------------|----------------|
| 1st<br>Declension | FEM<br>+Animate  | <i>dog</i>             | <i>sobak-a</i> | <i>sobak-u</i> | <i>sobak-i</i> |
|                   | FEM<br>-Animate  | <i>star</i>            | <i>zvezd-a</i> | <i>zvezd-u</i> | <i>zvezd-y</i> |
|                   | MASC<br>+Animate | <i>father</i>          | <i>pap-a</i>   | <i>pap-u</i>   | <i>pap-y</i>   |
| 2nd<br>Declension | MASC<br>-Animate | <i>table</i>           | <i>stol</i>    | <i>stol</i>    | <i>stol-a</i>  |
|                   | NEUT<br>-Animate | <i>window</i>          | <i>okn-o</i>   | <i>okn-o</i>   | <i>okn-a</i>   |
|                   | NEUT<br>-Animate | <i>butter</i>          | <i>masl-ǎ</i>  | <i>masl-ǎ</i>  | <i>masl-ǎ</i>  |

Unlike monolingual children who show almost error-free production of case morphology from early on, child HL-Russian speakers are reported to have difficulties with case morphology. In particular, this seems to be the case when their SL has sparse case morphology (like in English, Dutch, Swedish) and/or realizes cases differently, e.g., using particles (like in Hebrew) or marking case on determiners (like in German) (see Turian and Altenberg, 1991; Gagarina, 2011; Ringblom, 2012; Schwartz and Minkov, 2014; Janssen, 2016; Meir et al., 2017; Gagarina and Klassert, 2018; Janssen and Meir, 2019; Armon-Lotem et al., 2021). It should be noted that the studies differ with respect to the elicitation methods tapping into case morphology (e.g., *spontaneous speech*: Gagarina, 2011; Ringblom, 2012; Schwartz and Minkov, 2014; *experimental elicitation tasks*: Janssen, 2016; Janssen and Meir, 2019; Armon-Lotem et al., 2021, *sentence repetition tasks*: Meir et al., 2017), and with respect to the number of bilingual participants, ranging from case studies (Gagarina, 2011; Ringblom, 2012) to group comparisons (Janssen, 2016; Meir et al., 2017; Janssen and Meir, 2019; Armon-Lotem et al., 2021).

There is also evidence that the HL-Russian case system can be developed in a monolingual-like manner. For example, a study, investigating language development in a simultaneous bilingual child acquiring HL-Russian and SL-Turkish between 2;11 and 4;0, showed that the acquisition of case morphology is similar to that of monolingual children despite the reduced input of HL-Russian (Antonova Ünl and Li, 2018). It should be kept in mind that Turkish, the SL of the child, is an agglutinating language with rich case morphology. Turkish marks NOM, ACC case with definite noun phrases, DAT, LOC, GEN, and ablative case using dedicated case inflections. In the same vein, based on the data from an experimental task from 10 Russian-Finnish bilinguals aged 4;0–5;0, quantitative differences in HL-Russian in contact with Finnish were reported, but no qualitative differences were observed (Protassova et al., 2017). Russian-Finnish bilinguals rarely substituted various inflected case forms with the NOM default form.

To sum it up, while monolingual Russian-speaking children show almost error-free case production from early on, there is a conflict in the results for child HL-Russian speakers. On the one hand, some studies show monolingual-like error-free production in bilingual HL-Russian-speaking children. On the other hand, some studies point to profound difficulties with case morphology in bilingual HL-Russian-speaking children. There are several open questions that our study aims to answer. To what extent is case acquisition in the HL impeded/facilitated due to cross-linguistic influence from the SL, as has been proposed by previous studies? And to what extent do language-external factors mitigate possible negative cross-linguistic effects?

## PRESENT STUDY

The current study aimed to investigate bilingual HL development by evaluating the mechanisms which might shape morphosyntactic HL development. We considered the effects of cross-linguistic influence, i.e., influence from the SL,

and the effect of language-external factors, such as age, AoO of SL, home language use, family type, HL and SL input quantity, and HL and SL proficiency. Furthermore, we aimed to evaluate the interaction between these two potential mechanisms. Our specific research questions were as follows:

### RQ1: Is There an Effect of Cross-Linguistic Influence on Bilingual HL Development?

To address RQ1, we first employed bilingual vs. monolingual comparisons. Second, we compared the two bilingual groups between themselves (Russian-Dutch and Russian-Hebrew) to shed light on the cross-linguistic influence in HL acquisition i.e., the effect of the SL properties on the HL. Under the null hypothesis, we expected child HL speakers to perform on a par with their monolingual peers. The HL speakers were exposed to their HL from birth and acquired their HL *via* naturalistic input, thus, all groups might perform the same. We expected no group differences for the ACC and GEN forms which are homophonous to NOM, since, in both SLs (Dutch and Hebrew), the noun forms across different syntactic environments remain homophonous to the NOM form. Furthermore, under this hypothesis, no monolingual-bilingual differences in error patterns were expected.

Under the alternative hypothesis, we expected group differences. More specifically, group differences were predicted for the ACC and GEN forms which are non-homophonous to NOM, i.e., ACC and GEN forms that require the use of a dedicated inflection. Group differences and error patterns were expected to shed light on the mechanisms of HL development. If the presence and absence of the features in one language regulates the patterns of acquisition in the other language, Russian-Hebrew bilinguals were predicted to show better mastery of the ACC in HL-Russian as compared with the GEN case, as Hebrew marks the ACC case. Thus, under this scenario, the presence of ACC in Hebrew should facilitate the acquisition of ACC in HL-Russian. In the case of Russian-Dutch bilinguals, such asymmetry was not expected, as Dutch does not have any morphological markings for any of these environments. Alternatively, based on the *feature (re)-assembly hypothesis* (Lardiere, 2009), we would expect no asymmetry between ACC and GEN in the bilingual groups, since the ACC case is mapped onto different lexical categories in the two languages of Russian-Hebrew bilinguals (onto case inflection in Russian and particle *et* in Hebrew), thus, these differences in realization are expected to impede the acquisition/maintenance of ACC in Russian-Hebrew bilinguals. Similarly, no asymmetry was expected in HL-Russian among Dutch bilinguals. Finally, if cross-linguistic influence shapes HL acquisition, we expected qualitative differences between monolinguals and bilinguals.

### RQ2: Do Language-External Factors Affect Bilingual HL Development?

To address RQ2, we aimed to evaluate whether age, AoO of SL, home language use, family type, HL and SL input quantity, and HL and SL proficiency affects the acquisition of case morphology in bilingual HL-Russian development. Previous research brings robust evidence for the key role of language-external factors for

the acquisition of vocabulary in HL and SL, yet, with respect to the acquisition of HL morphosyntactic properties, the findings are less robust. Some studies show that various language-external factors might shape morphosyntactic HL development, while other studies fail to detect the effect of language-external factors (e.g., Daskalaki et al., 2019, 2020; Rodina et al., 2020; Armon-Lotem et al., 2021; Otwinowska et al., 2021).

In the current study, we did not consider qualitative input factors, although previous studies reported parental HL input divergences from the one provided by monolingual baseline speakers (e.g., Daskalaki et al., 2020). Immigrants constantly mix language in the same utterance, including Russian-speaking immigrants (e.g., Remennick, 2003). However, the Russian-speaking immigrants of the first generation inflect for case code-switched lexical items inserted into a Russian sentence, following the Russian system of case assignment based on declension classes. The correct assignment of cases (even on code-switched words) is indicative of intact grammatical structure in Russian. Furthermore, two recent studies showed no traces of divergences in case production and comprehension in Russian among first-generation immigrants (see Meir and Polinsky, 2021). Thus, based on previous studies, input type divergences were not expected to account for the ACC and GEN case differences in child HL.

### RQ3: Do Language-External Factors Mitigate the Potential Effects of Cross-Linguistic Influence?

To contribute to our understanding of the interplay between cross-linguistic influence and language-external factors, we assessed whether language-external factors mitigate the effects of cross-linguistic influence. If language-external factors mitigate cross-linguistic influence effects, we expected interactions between language-external factors and group membership (biDU or biHE). The interactions between language-external factors and groups membership were expected to show the differential effects of SL properties across the two cases, ACC and GEN, given the inherent differences of Dutch and Hebrew. The presence of an interaction between language-external factors and cross-linguistic inputs would support the claim that input is not the only factor shaping HL development. The interaction between the language-external and cross-linguistic influence would point to an interplay between these two factors suggesting that HL development is a complex process influenced by multiple mechanisms.

## METHODOLOGY

### Participants

Three groups of children and a group of monolingual adult controls participated in the study. The adult monolingual Russian-speaking controls ( $n = 10$ ) were recruited from the Russian Federation. The adult controls were between the ages of 18 and 45, residing in Saint Petersburg. The monolingual child controls (monoRU) were recruited from the Russian Federation, while the bilingual HL-Russian speakers were residing in

the Netherlands (biDU) or Israel (biHE). The background information on the child data is presented in **Table 2**. The current sample ( $n = 115$ ) partially overlaps with the sample of children ( $n = 72$ ) reported by Janssen and Meir (2019) that compared the accuracy production, comprehension, and repetition of the ACC case forms. In the current study, we investigated two cases: ACC and GEN.

There were significant group differences for age [ $F_{(2,112)} = 3.67$ ,  $p = 0.030$ ,  $\eta^2 = 0.062$ ]: the biDU group was significantly older than the monoRU group ( $p = 0.031$ ) as determined by the Bonferroni *post-hoc* analysis for pair-wise comparisons; no differences for age were detected between the biHE and the monoRU ( $p = 1.00$ ), and between the two bilingual groups ( $p = 0.17$ ). Similarly, there were significant differences between the groups with respect to the level of education of their mothers [ $\chi_{(1)} = 14.99$ ,  $p = 0.005$ ]: the percentage of mothers holding a university degree was lower in the biHE group.

The biDU and biHE groups differed in the length of uninterrupted HL acquisition (i.e., AoO) as determined by Welch's *t*-test for unequal variance [ $t_{(52,21)} = 5.32$ ,  $p < 0.001$ ]: the children in the biHE had far later SL AoOs. In turn, there were significant differences for the length of exposure to SL, which was calculated as the difference between the chronological age and AoO: a Welch's *t*-test for unequal variance showed [ $t_{(61,13)} = 6.46$ ,  $p < 0.001$ ]. It should be noted that the AoO and LoE were highly correlated in the current sample of children [ $r_{(74)} = 0.90$ ,  $p < 0.001$ ]. There were differences between the two groups concerning the amount of current exposure to HL-Russian [ $t_{(69)} = 4.57$ ,  $p < 0.001$ ]. The bilingual groups differed with respect to family type (HL speaking family vs. mixed family) [ $\chi_{(1)} = 13.07$ ,  $p < 0.001$ ]: in the biDU sample, HL families amounted to 35% of the sample, while in biHE they amounted to 77% of the sample. In the biHE group, 54.3% of the families reported Russian to be the language of communication; while in the biDU group, only 33.3% did; this difference did not reach significance [ $\chi_{(1)} = 3.30$ ,  $p = 0.07$ ].

In the biDU sample, there were no differences in the parental ratings for HL and SL proficiency (i.e., *How would you rate the language proficiency of your child in HL/SL on the scale of 1–4*), as determined by the paired *t*-test [ $t_{(36)} = 0.14$ ,  $p = 0.87$ ], suggesting that the parents viewed their children as balanced bilinguals. Alternatively, in the biHE group, there were significant differences between HL and SL proficiency [ $t_{(33)} = 3.24$ ,  $p = 0.003$ ], thus indicating that the parents reported the HL proficiency of the biHE children as higher compared with the SL proficiency.

To sum it up, the background differences between the two bilingual groups reflect the nature of the Russian-speaking populations in the Netherlands and Israel. These inherent background differences will be included in the analysis to understand which factors are related to the success/failure of case morphology acquisition in bilingual children.

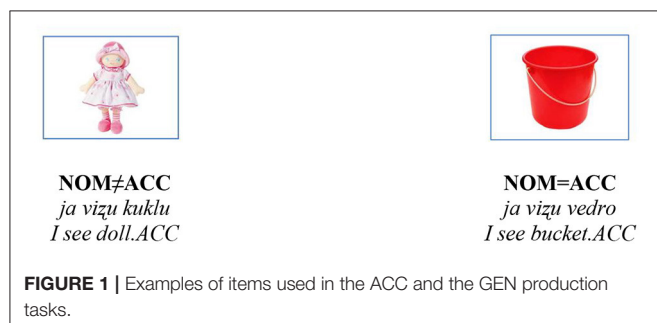
### Tasks

#### The Accusative Case Elicitation Task

The ACC case elicitation task investigates the accuracy of ACC inflection production in 36 nouns (Janssen, 2016; Janssen

**TABLE 2 |** Background information on the participants per group.

|                                                             |                                 | biDU (n = 39)                              | biHE (n = 35)                              | monoRU (n = 41)    |
|-------------------------------------------------------------|---------------------------------|--------------------------------------------|--------------------------------------------|--------------------|
| Age (months)                                                | <i>M (SD)</i><br><i>MIN-MAX</i> | 61 (8)<br>48–77                            | 57 (9)<br>43–70                            | 56 (8)<br>41–71    |
| Length of uninterrupted HL acquisition (as measured by AoO) | <i>M (SD)</i><br><i>MIN-MAX</i> | 7 (11)<br>0–39                             | 27 (20)<br>0–48                            | n/a                |
| Length of exposure (months) to the Societal Language (SL)   | <i>M (SD)</i><br><i>MIN-MAX</i> | 54 (13)<br>26–77                           | 30 (18)<br>0–69                            | n/a                |
| Mother's Education (% of mothers with a university degree)  | %                               | 95%                                        | 63%                                        | 83%                |
| Language spoken at home                                     |                                 | Only HL-Russian: 33.3%<br>HL and SL: 66.7% | Only HL-Russian: 54.3%<br>HL and SL: 45.7% | Only Russian: 100% |
| Parental ratings of the HL-Russian skills                   | <i>M (SD)</i><br><i>MIN-MAX</i> | 2.78 (0.82)<br>1–4                         | 2.81 (0.70)<br>1–4                         | n/a                |
| Parental ratings of the SL skills                           | <i>M (SD)</i><br><i>MIN-MAX</i> | 3.35 (0.77)<br>1–4                         | 2.46 (1.07)<br>1–4                         | n/a                |
| Current exposure to the HL-Russian (0–1)                    | <i>M (SD)</i><br><i>MIN-MAX</i> | 0.35 (0.18)<br>(0–0.75)                    | 0.58 (0.24)<br>(0.25–1.00)                 | n/a                |
| Family type                                                 | %                               | HL: 35%<br>mixed: 65%                      | HL: 77%<br>Mixed: 23%                      | n/a                |



and Meir, 2019). The nouns were all 2- or 3-syllable words, most of the items were taken from the Russian MacArthur-Bates Communicative Development Inventories (Vershina and Eliseeva, 2007), thus they were all very frequent in child-directed speech, and they were all familiar to young Russian-speaking children. For this task, the pictures were grouped thematically (in groups of four or five items). For example, 3–4 pictures of fruit items (e.g., apple, strawberry, pear) and 3–4 food items (e.g., milk, sausage, egg). The child was asked to describe what he/she sees on the card by saying *ja vizu* \_\_\_\_ “I see (target noun).” If the child failed to respond to the sentence with *ja vizu* \_\_\_\_ “I see \_\_\_\_,” he was reminded to start the sentence with *ja vizu*. “I see.” This was done for each target noun to ensure that the syntactic environment for the ACC case was produced. The task included 12 feminine items, 12 masculine items, and 12 neutral items which varied with respect to whether the ACC form was homophonous or not to the NOM form. See **Figure 1**.

When the target ACC inflection was produced, the response was coded as “correct” and one point was given. Responses with non-target inflections were coded as “incorrect” and zero points were given. In addition, we noted the type of error in the ACC condition (the use of NOM singular, NOM plural,

over-generalization/over-extension of inflections, -u (\**myl-u*), -ov (\**knig-ov*), -i/y (\**myl-y*), -a (*arbuza*).

### The Genitive Case Elicitation Task

The GEN case elicitation task (Janssen, 2016) examined the production of GEN in the context of negation. The child was asked to name the picture *vo* \_\_\_\_ “here is (target noun),” then to turn a picture over and say *a teper net* \_\_\_\_ “and now there is no noun. GEN.” The task included the same 36 items as the ACC case task (see **Figure 1**). Similar to the ACC task, some nouns in GEN forms are homophonous to the NOM (e.g., *mjaso* “meat”, *jabloko* “apple”).

The responses of the children were coded as “correct” when the target GEN inflection was produced. Responses with non-target inflections were coded as ‘incorrect’. In addition, similar to the ACC task, we noted the type of error in the GEN condition -u (\**myl-u*), -ov (\**knig-ov*), -i/y (\**myl-y*).

### Procedure

Informed parental consent was secured for each child before the testing session, as well as the oral ascent of the child. Each participant was tested individually. The experimenter gave oral instructions. Four warm-up items were administered to familiarize the children with the task: the trial items were not included in the analysis. The study was approved by the Ethics Committee of the University of Amsterdam and the review board of Bar-Ilan University.

The administration of both tasks took an average of around 5 min to complete. The responses of the children were audio-recorded and then transcribed and coded off-line.

### RESULTS

The descriptive results for the accuracy of the ACC and GEN production are presented in **Figure 2** for the child and adult data. The results indicate that monolingual adults show

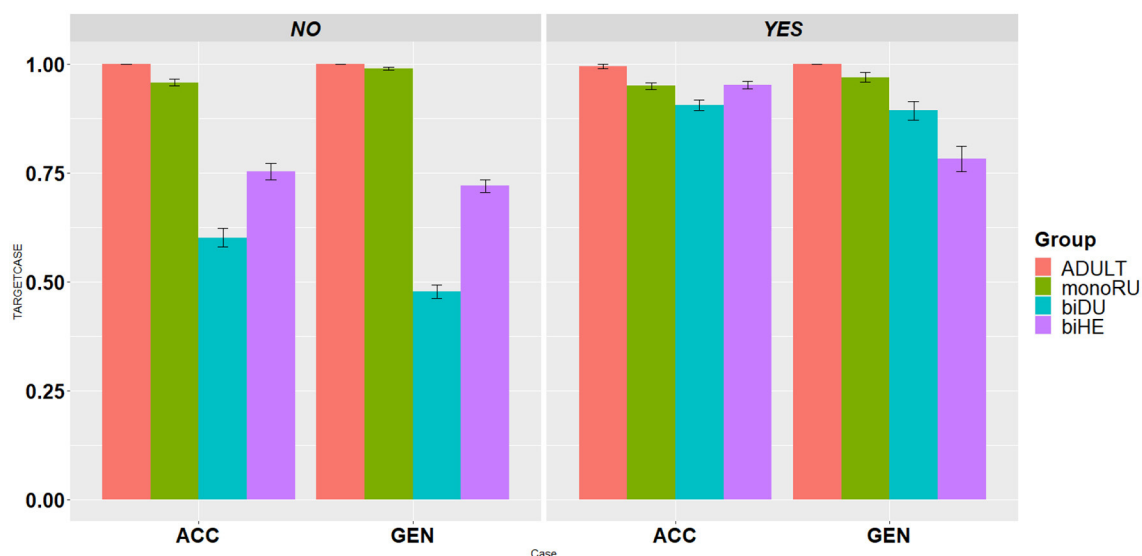


**FIGURE 2 |** Mean accuracy scores (with standard error bars) per case (ACC vs. NOM) per group (ADULT, monoRU, biDU, biHE) per inflection type (Homophonous to NOM: NO vs. YES).

a ceiling performance, confirming our target vs. non-target coding. Similarly, Russian-speaking monolinguals showed a near-ceiling performance (above 0.95) for both ACC and GEN forms. Bilinguals (both biDU and biHE) showed a near-ceiling performance on the forms which are homophonous to NOM in the ACC and GEN contexts, while on the non-homophonous forms which require the use of a dedicated inflection, the performance was lower.

## Effects of Cross-Linguistic Influence in HL Morpho-Syntactic Acquisition

Following previous studies, our first statistical analysis aimed to assess the potential effects of SL properties on HL acquisition by comparing bilinguals and monolinguals.

Given the binary nature of the data (Target/Non-Target), we analyzed the child data using a mixed-effects binomial regression with the group (monoRU, biDU, biHE), case (ACC vs. GEN), inflection type (homophonous to NOM: no vs. yes) as fixed factors. The models were built by adding random and fixed variables in a step-by-step procedure, starting with an intercept-only baseline model. The null models included both by-subject random intercepts and by-stimulus random intercepts. With the inclusion of random slopes, the models failed to converge, and therefore random slopes were not included in the final models. First, we built the model starting with the group, and then we added the case and inflection type as fixed factors. The variables and/or the interactions of the variables were retained in the model only if they significantly improved the fit of the model, resulting in a reduced Akaike information criterion (AIC)-value. We also included a three-way interaction Group\*Case\* Inflection Type. The inclusion

of two-way interactions did not significantly improve the fit of the model. The comparison of the models was carried out using one-way ANOVAs, the *p*-value which is higher than 0.05 indicated that the variable and/or the interaction does not improve the goodness of the fit. The analysis was conducted using R (R Core Team, 2020). The final minimal adequate model performed significantly better than the minimal baseline model.

The model summary is presented in **Table 3**. The results indicated a significant effect on the groups: both bilingual groups were significantly different from their monolingual peers. There was a significant effect on the Case (ACC vs. GEN), and there were significant interactions (Group\*Inflection Type, Group\*Case and Group\*Case\*Inflection Type).

As a follow-up on the significant three-way interaction (see **Table 4**), pair-wise comparisons with an adjusted alpha level using the Tukey method showed that on the non-homophonous forms, i.e., the forms that require the use of a dedicated inflection to mark ACC/GEN cases, both bilingual groups were significantly less accurate as compared with their monolingual peers. As for the forms, which are homophonous to NOM, the biDU did not differ from monolingual controls, while the biHE differed only on the GEN forms. Thus, quantitative differences emerged between monolinguals and the two bilingual groups on nouns that require dedicated inflections in ACC and GEN forms which are different from NOM ones.

To shed further light on the contribution of SL properties to HL acquisition, we compared the accuracy of production of ACC and GEN cases in the three child groups (monoRU, biDU, and biHE). The adjusted pair-wise comparisons using the Tukey method indicated that while monolingual controls were more accurate on the GEN forms, the two bilingual groups were more

**TABLE 3 |** Model Summary for target case production for the child data.

|                                                   | Estimate    | SE     | Z      | Sig.    |
|---------------------------------------------------|-------------|--------|--------|---------|
| (Intercept)                                       | 3.4882      | 0.317  | 11.003 | <0.0001 |
| Group (monoRU vs. biDU)                           | −3.1096     | 0.3828 | −8.124 | <0.0001 |
| Group biHE (monoRU vs. monoRU)                    | −2.1421     | 0.3941 | −5.435 | <0.0001 |
| Case (ACC vs. GEN)                                | 1.6972      | 0.3372 | 5.033  | <0.0001 |
| Inflection_TYPE (No vs. YES)                      | 0.2603      | 0.2755 | 0.945  | 0.345   |
| Group (biDU) * Inflection TYPE (YES)              | 2.4342      | 0.3119 | 7.804  | <0.0001 |
| Group (biHE) * Inflection TYPE (YES)              | 2.5873      | 0.3481 | 7.433  | <0.0001 |
| Group (biDU) * Case (GEN)                         | −2.2181     | 0.3567 | −6.219 | <0.0001 |
| Group (biHE) * Case (GEN)                         | −1.6742     | 0.3636 | −4.605 | <0.0001 |
| Case (GEN) * Inflection TYPE (YES)                | −1.5356     | 0.5422 | −2.832 | <0.001  |
| Group (biDU) * Case (GEN) * Inflection TYPE (YES) | 1.562       | 0.6022 | 2.594  | <0.001  |
| Group (biHE) * Case (GEN) * Inflection TYPE (YES) | −1.0611     | 0.6114 | −1.736 | 0.082   |
| Observations                                      | 7,513       |        |        |         |
| Marginal R2/Conditional R2                        | 0.379/0.618 |        |        |         |

**TABLE 4 |** Pairwise contrasts for group comparisons (monoRU vs. biDU, monoRU vs. biHE) per case per inflection type.

| Case | Inflection type (homophonous to NOM) | Group pairwise contrasts | Contrast estimate | SE    | Z      | Adj. Sig. |
|------|--------------------------------------|--------------------------|-------------------|-------|--------|-----------|
| GEN  | NO                                   | <b>monoRU—biDU</b>       | 5.328             | 0.419 | 12.724 | <0.0001   |
|      |                                      | <b>monoRU—biHE</b>       | 3.816             | 0.427 | 8.928  | <0.0001   |
|      | YES                                  | monoRU—biDU              | 1.332             | 0.536 | 2.485  | 0.350     |
|      |                                      | monoRU—biHE              | 2.290             | 0.527 | 4.347  | 0.0008    |
| ACC  | NO                                   | <b>monoRU—biDU</b>       | 3.110             | 0.383 | 8.124  | <0.0001   |
|      |                                      | <b>monoRU—biHE</b>       | 2.142             | 0.394 | 5.435  | <0.0001   |
|      | YES                                  | monoRU—biDU              | 0.675             | 0.381 | 1.773  | 0.833     |
|      |                                      | monoRU—biHE              | −0.445            | 0.415 | −1.073 | 0.996     |

**TABLE 5 |** Pair-wise ACC-vs-GEN contrasts for monoRU, biDU and biHE groups.

| Group  | Contrast estimate | SE    | Z      | Adj. Sig. |
|--------|-------------------|-------|--------|-----------|
| monoRU | −0.929            | 0.269 | −3.455 | 0.0073    |
| biDU   | 0.508             | 0.163 | 3.110  | 0.0230    |
| biHE   | 1.275             | 0.173 | 7.386  | <0.0001   |

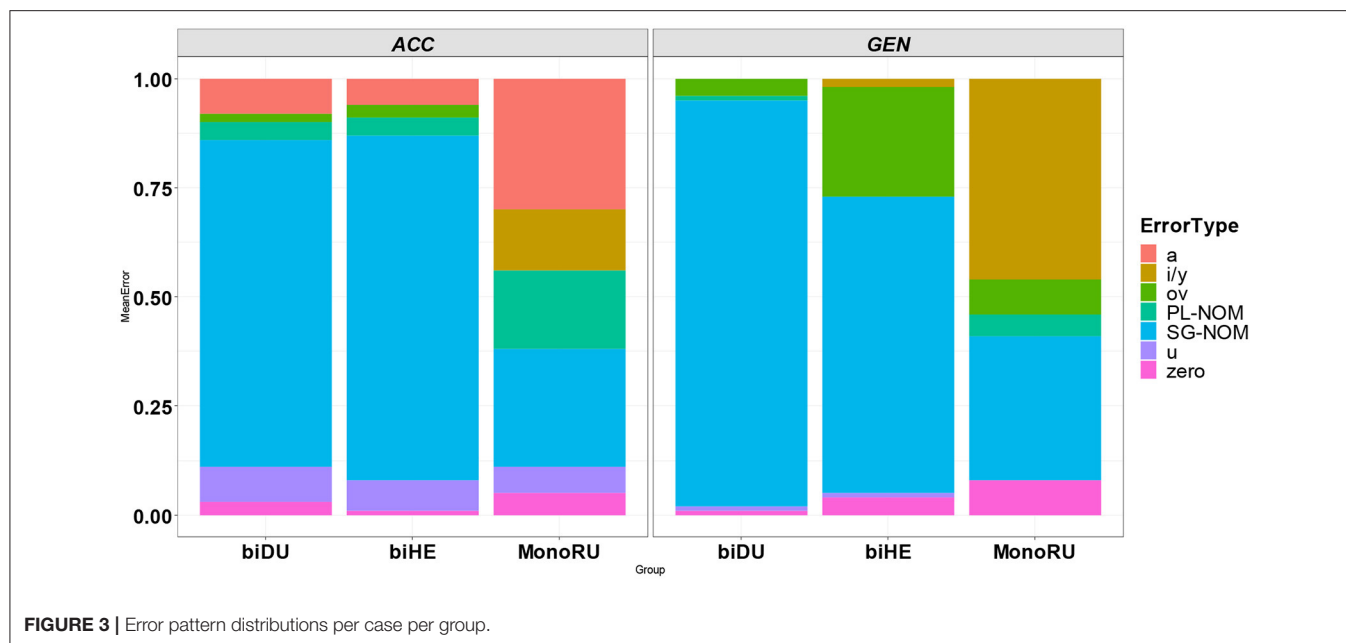
accurate on the ACC forms as compared with the GEN ones (see **Table 5**). It is worth noticing that the gap between ACC and GEN was more likely to be observed in the biHE group.

Subsequently, we conducted a qualitative analysis of error patterns across the three child groups to shed light on whether monolingual-bilingual group differences were only quantitative or also qualitative (see **Figure 3**), i.e., whether bilingual and monolingual children resorted to different types of errors. Indeed, the error pattern analysis revealed that there are qualitative differences: the most common error pattern among bilingual children was the use of default NOM forms in ACC and GEN contexts (biDU: ACC = 0.75; GEN = 0.92; biHE: ACC = 0.79; GEN = 0.68). In the monolinguals, errors were very infrequent totaling 87 (63 in the ACC contexts and 24 in the GEN contexts): the use of NOM forms was not the

most prevalent type of error (ACC = 0.27; GEN = 0.33). The erroneous substitution of other non-target inflections was more common in the monoRU group leading to innovations (e.g., ACC: *a teper' net \*promidory / \*myly / \*mjasy*; GEN: *ja vižu \*jabloku / \*jajcu*).

In contrast, bilingual children resorted to NOM forms in both ACC contexts (*ja vižu \*kukla / \*klubnika / \*krokodil / \*butylka* 'I see doll.NOM' / strawberry.NOM / crocodile.NOM / bottle.NOM), as well as in GEN contexts (*a teper' net \*gruša / \*karandas / \*kolbasa* 'and there is no pear.NOM / pencil.NOM / sausage.NOM').

To sum up, the results of the first analysis, which compared monolinguals vs. bilinguals, showed that bilingual children were less accurate than their monolingual peers on nouns that require the use of dedicated ACC and GEN inflections (i.e., non-homophonous inflections). Furthermore, unlike monolingual child controls who were more accurate on the GEN forms as compared with the ACC forms, the picture was reversed in the bilingual groups. We predicted the gap to be observed only in the biHE group, due to the inherent properties of SL-Hebrew which mark the ACC case, but not in the biDU group. It is important to note that the gap between ACC and GEN was more likely to be detected in the biHE group. In addition, both bilingual groups showed different error profiles as compared with



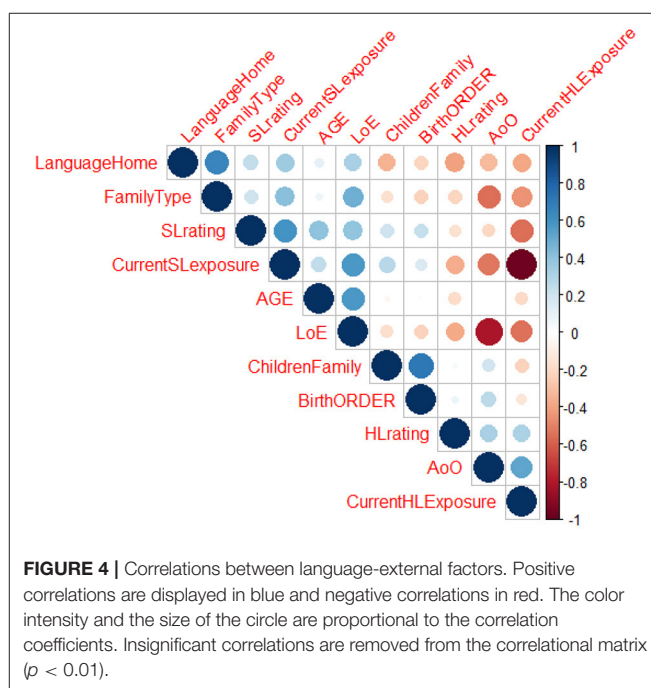
monolinguals. While monolinguals substituted the target forms with other non-target inflected forms, bilinguals favored the use of NOM forms.

In our subsequent analyses, we focused on the two bilingual groups. First, we evaluated the contribution of language-external variables in bilingual HL development. Second, we attempted to assess the contribution of cross-linguistic influence and language-external factors.

### The Effects of Language-External Factors

To address our second research question on the role of language-external factors in bilingual HL development, we first looked at the interrelationship between various language-external factors. This was done to determine which factors should be included in the subsequent statistical models. The results indicated medium to strong correlations between different language-external factors (see **Figure 4**): age, AoO, length of exposure to SL (LoE), language spoken at home, family type, current exposure to HL and SL, parental rating of child HL and SL proficiency, number of children in the family, and first-born or not. For example, the results indicated that there were strong negative correlations between the current HL and SL exposure indices ( $r = -0.98$ ), suggesting that children with increased current exposure to HL have less exposure to SL. There were strong negative correlations between AoO and LoE ( $r = -0.81$ ): children with earlier AoO had more exposure to L2. Thus, following the observed correlations between some of the language-external factors, it was decided to exclude some variables from further analyses (if the correlation between the two variables was above 0.7).

To address the more specific aspect of our second research question concerning the role of language-external factors in HL morphosyntactic acquisition, as in the previous model we started with case production accuracy as the outcome variable coded in a binary manner (Target/Non-Target), and random factors



(Participant and Item). We included case and inflection type, as well as language-external variables. The best model included case and inflection type, plus AoO and HL parental rating and the interactions between inflection type and AoO, and inflection type and HL parental rating (see **Table 6**). We note that chronological age, type of family, current exposure to HL, number of children in the family, and the status of being firstborn did not improve the fit of the model, therefore these variables were excluded from the final model.

**TABLE 6 |** Model summary for target case production and language-external factors.

|                                      | Estimate    | SE    | Z      | Sig.    |
|--------------------------------------|-------------|-------|--------|---------|
| (Intercept)                          | −2.655      | 0.681 | −3.896 | <0.0001 |
| Inflection Type (yes/no)             | 6.146       | 0.487 | 12.615 | <0.0001 |
| Case (GEN vs. ACC)                   | −0.551      | 0.104 | −5.274 | <0.0001 |
| AoO                                  | 0.029       | 0.010 | 2.814  | 0.006   |
| HL rating                            | 1.031       | 0.230 | 4.482  | <0.0001 |
| Inflection Type (yes/no) * AoO       | −0.013      | 0.007 | −1.964 | 0.049   |
| Inflection Type (yes/no) * HL rating | −1.194      | 0.154 | −7.757 | <0.0001 |
| Observations                         | 4,383       |       |        |         |
| Marginal R2/Conditional R2           | 0.304/0.566 |       |        |         |

The results of the model are illustrated in **Figures 5A,B**. Language-external factors affect the case accuracy production of nouns which require the use of dedicated case inflections only. While language-external factors do not affect forms that are homophonous to NOM, the accuracy of the case production (which requires the use of a dedicated inflection, ACC and/or GEN) is predicted by AoO and HL ratings. Children with later AoOs of bilingualism are more likely to produce target case inflections (see **Figure 5A**). The same applies to HL rating: with the growing proficiency in HL, the accuracy of the target case inflections improves (see **Figure 5B**). To summarize, the production of target case inflections increases with later AoOs of bilingualism and growing proficiency in HL.

## The Effects of Cross-Linguistic Influence and Language-External Factors

Finally, to address our third research question which aimed to evaluate the interplay between cross-linguistic influence and language-external factors in bilingual children, we further compared the two bilingual groups. These two groups (biDU and biHE) varied with respect to the case properties in the SL. While the Dutch language does not mark cases morphologically on nouns, the Hebrew language has a dedicated particle “*et*” to mark the ACC case. Thus, the purpose of our third model was to evaluate the contributions of SL factors and language-external factors to HL morphosyntactic acquisition. Based on the two previous models, we fitted a model with case production accuracy as the outcome variable, and participant and item as random factors. We attempted to assess the presence of three-way interactions between case (ACC vs. GEN), AoO, and group (biDU vs. biHE), on one hand, and between case (ACC vs. GEN), HL Rating, and group (biDU vs. biHE), on the other hand. While the three-way Group\*Case\*AoO interaction did not improve the fit of the model, the Group\*Case\*HL rating did significantly improve it. The estimates of the final model are presented in **Table 7**.

The three-way interaction indicated that there is an interplay between language-external factors and cross-linguistic influence. The growth in the biHE group is steeper for the ACC case as compared with the biDU group which might indicate the positive direct effect of ACC marking in the Hebrew language.

Furthermore, there is also a steep growth for the GEN case in the biHE group, which might indicate an indirect positive effect from the SL that marks cases morphologically. In the biDU, the increase in HL proficiency is associated with the parallel growth for ACC and GEN, with no advantage for ACC (see **Figure 6**).

## DISCUSSION

The current study aimed to investigate the potential mechanisms affecting child HL development. HL speakers are exposed to their HL from birth *via* naturalistic input. Adult HL speakers were reported to show divergences in a number of language domains with morphosyntax being the hallmark of difficulties in HL adult grammars (for an overview see Montrul, 2016; Polinsky, 2018). The current study employed monolingual vs. bilingual comparisons, as well as between-bilingual-group comparisons to contribute to the knowledge on the development of child HL grammar. Most of the previous studies investigating cross-linguistic influence compared the performance of bilinguals with that of monolinguals; fewer carried out multiple bilingual group comparisons. Van Dijk's (2021) meta-analysis investigating cross-linguistic influence via bilingual vs. monolingual comparisons confirmed the presence of cross-linguistic influence in the acquisition of morphosyntax based on 26 studies. However, some studies employing between-bilingual-group comparisons did not find SL influence on HL and point to the effect of input characteristics (see the introductory section of this paper). In line with previous studies, our results revealed a complex picture of HL development.

Firstly, the results of the current study showed that the domain of morphosyntax is not only challenging for adult HL-Russian speakers (see Polinsky, 2018 and references in it), but also for child HL-Russian speakers. Monolingual Russian-speaking children showed a near-ceiling performance in line with the previous literature (e.g., Gvozdev, 1961; Babyonyshev, 1993; Protassova and Voeikova, 2007; Cejtin, 2009). In contrast, both bilingual groups were less accurate in case morphology compared to their monolingual peers. This has been shown for the ACC case as well as the GEN case. We also found different error patterns in the bilingual groups. While monolingual Russian-speaking controls produced non-target inflected forms, bilingual HL-Russian speakers resorted to the default NOM forms. The overuse of NOM was the most common error in the bilingual groups. The results of the current study corroborate previous findings for child HL-Russian speakers (see Turian and Altenberg, 1991; Gagarina, 2011; Ringblom, 2012; Schwartz and Minkov, 2014; Janssen, 2016; Meir et al., 2017; Protassova et al., 2017; Gagarina and Klassert, 2018; Janssen and Meir, 2019; Armon-Lotem et al., 2021). Most of the previous studies concluded that the reduced accuracy in case morphology is driven by the properties of the SLs, as the SLs do not mark case morphology in nouns. This is indeed a plausible explanation, and our study does not rule out this option.

To deepen our understanding of the mechanisms of HL development, we conducted a comparison between the two bilingual groups, beyond the monolingual vs. bilingual

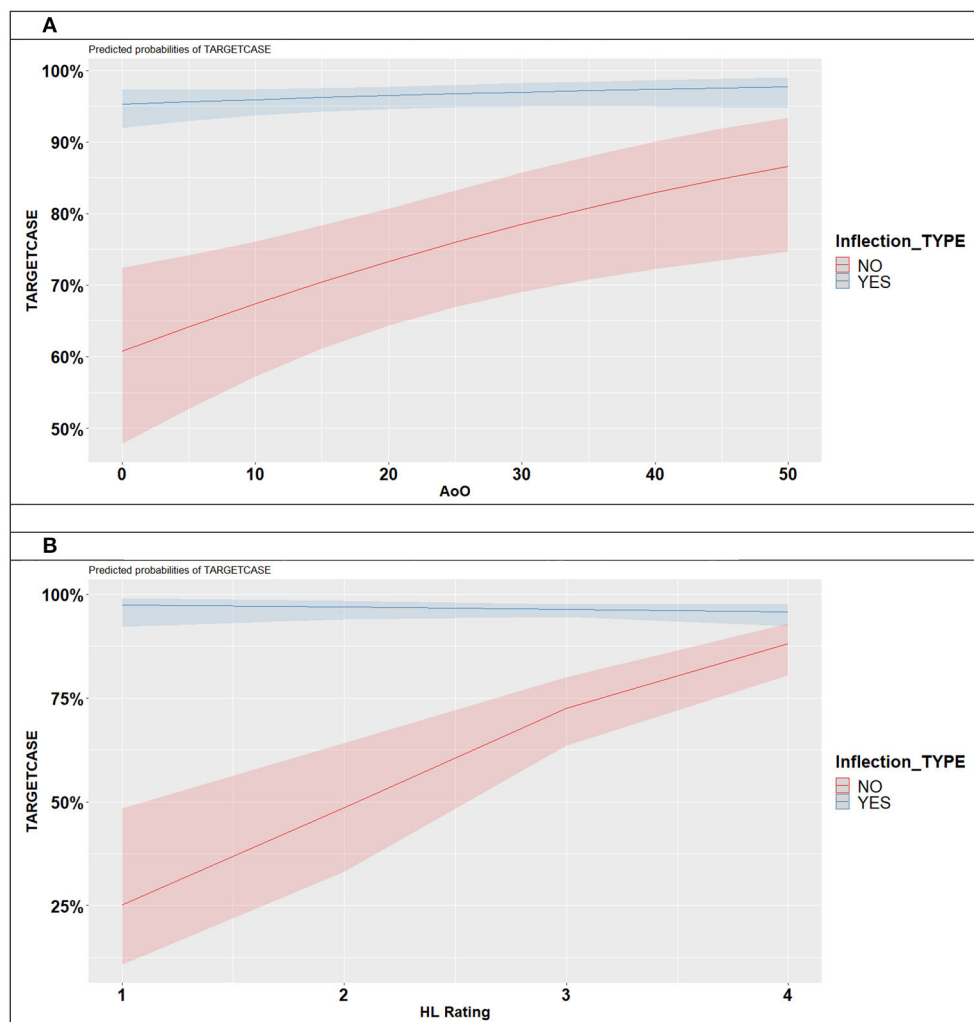


**FIGURE 5 | (A)** Predicted case production by inflection type by AoO. **(B)** Predicted case production by inflection type by HL rating.

comparisons. Studies comparing different bilingual groups matched on one language are scarce. The two bilingual groups in this study differed with respect to the SL properties. Dutch does not mark cases morphologically on nouns. Hebrew marks ACC case with the particle “*et*” in front of definite nouns. Thus, we hypothesized that if the presence/absence of a feature in the language affects HL acquisition, Russian-Hebrew bilinguals would show better mastery of the ACC case as compared with the GEN case in HL-Russian. No ACC-GEN asymmetry was expected in the bilingual Russian-Dutch group if the properties of SL shape HL acquisition. Alternatively, under the *feature (re)-assembly hypothesis* (Lardiere, 2009), we predicted no facilitative effect for the acquisition of ACC in Russian-Hebrew bilinguals, since the ACC case is mapped onto different lexical categories in the two languages of Russian-Hebrew bilinguals (onto case inflection in Russian and particle *et* in Hebrew). Under this scenario, no asymmetry was expected between the ACC and GEN case production. Similar to previous studies, multiple bilingual group comparisons yielded complex

results. First, the comparison of the case production in HL-Russian among Russian-Dutch and Russian-Hebrew bilinguals showed an advantage for the ACC forms in both groups, while in the monolingual controls, the reverse picture was observed, with better performance on GEN forms. Based on the frequency of ACC and GEN forms in the general corpus, no differences were expected. However, our study did not include the frequencies of ACC and GEN form in the child-directed speech. Future studies need to include frequencies of forms in child-directed speech to further shed light on the mechanisms of HL development, thus focusing more on quantitative and qualitative input characteristics.

Language-external variables play an important role in monolingual language acquisition and even more so in bilinguals (see Armon-Lotem and Meir, 2019). Thus, it is plausible to suggest that monolingual vs. bilingual differences (also observed in the current study) are not solely driven by the properties of the SL, but also related to language-external variables. Indeed, recent studies comparing multiple bilingual groups

showed that it is not cross-linguistic influence that shapes HL development, but rather, for the most part, language-external variables (see Rodina et al., 2020). The current study confirms that language-external variables are an important part of the puzzle of bilingual HL development. The findings show that the AoO of bilingualism and HL proficiency (as indexed by HL parental ratings) affect the HL morphosyntactic development. Longer periods of uninterrupted acquisition of HL and growing HL proficiency contribute to the success of the HL morphosyntactic development, and as a result language-external factors compensate for the possible negative effects of the SL on the HL.

**TABLE 7 |** Model summary for target case production (cross-linguistic influence and language-external variables).

|                              | Estimate    | SE    | Z      | Sig.    |
|------------------------------|-------------|-------|--------|---------|
| (Intercept)                  | −1.384      | 0.835 | −1.657 | 0.098   |
| Inflection Type (yes vs. no) | 2.452       | 0.170 | 14.483 | <0.0001 |
| Case (GEN vs. ACC)           | −0.413      | 0.421 | −0.982 | 0.326   |
| Group                        | −0.667      | 1.594 | −0.418 | 0.676   |
| HL rating                    | 0.668       | 0.289 | 2.312  | 0.021   |
| Case (GEN) * Group (biHE)    | −1.970      | 0.848 | −2.322 | 0.020   |
| Group (biHE) * HL rating     | 0.378       | 0.489 | 0.774  | 0.439   |
| Case (GEN) * HL rating       | −0.076      | 0.151 | −0.500 | 0.617   |
| Case (GEN) * Group*HL rating | 0.686       | 0.267 | 2.575  | 0.010   |
| Observations                 | 4,383       |       |        |         |
| Marginal R2/Conditional R2   | 0.330/0.577 |       |        |         |

Interestingly, it was AoO, rather than chronological age, that predicted case production in the HL. The effect of chronological age is robust in the monolingual acquisition, yet not so strong in bilingual HL and SL morphosyntactic development. Our findings showed that chronological age had no effect on HL morphosyntactic development, in line with the results of the meta-analysis by Van Dijk et al. (2021) which showed that age does not module the extent of cross-linguistic influence. However, AoO has been shown to be one of the key factors shaping HL development in bilinguals in previous studies investigating SL and HL morphosyntactic development (see Tsimpli, 2014; Meir et al., 2017; Armon-Lotem et al., 2021). In addition, we found that HL proficiency is a powerful predictor of case production in HL. As the HL proficiency increases, children are more likely to have a target-like case system. This has also been related to HL and SL input/exposure factors. As previously noted, as HL speakers grow older, they switch from their HL to the SL, which is visible in the increased SL input/exposure and growing SL proficiency, while there is a simultaneous decrease in HL input exposure. Indeed, our correlational analyses showed that HL proficiency was related AoO, current HL exposure, and family type. All these variables contribute to a higher level of HL proficiency (see Figure 4).

Furthermore, our study attempted to evaluate the interplay between cross-linguistic variables and language-external factors, i.e., trying to evaluate the mitigators of cross-linguistic influence. The meta-analysis by Van Dijk et al. (2021) investigating language dominance as a possible moderator of cross-linguistic influence, showed that the effect of the SL (the dominant language) on the HL (the non-dominant language) is stronger compared with the influence in the opposite direction. The authors showed

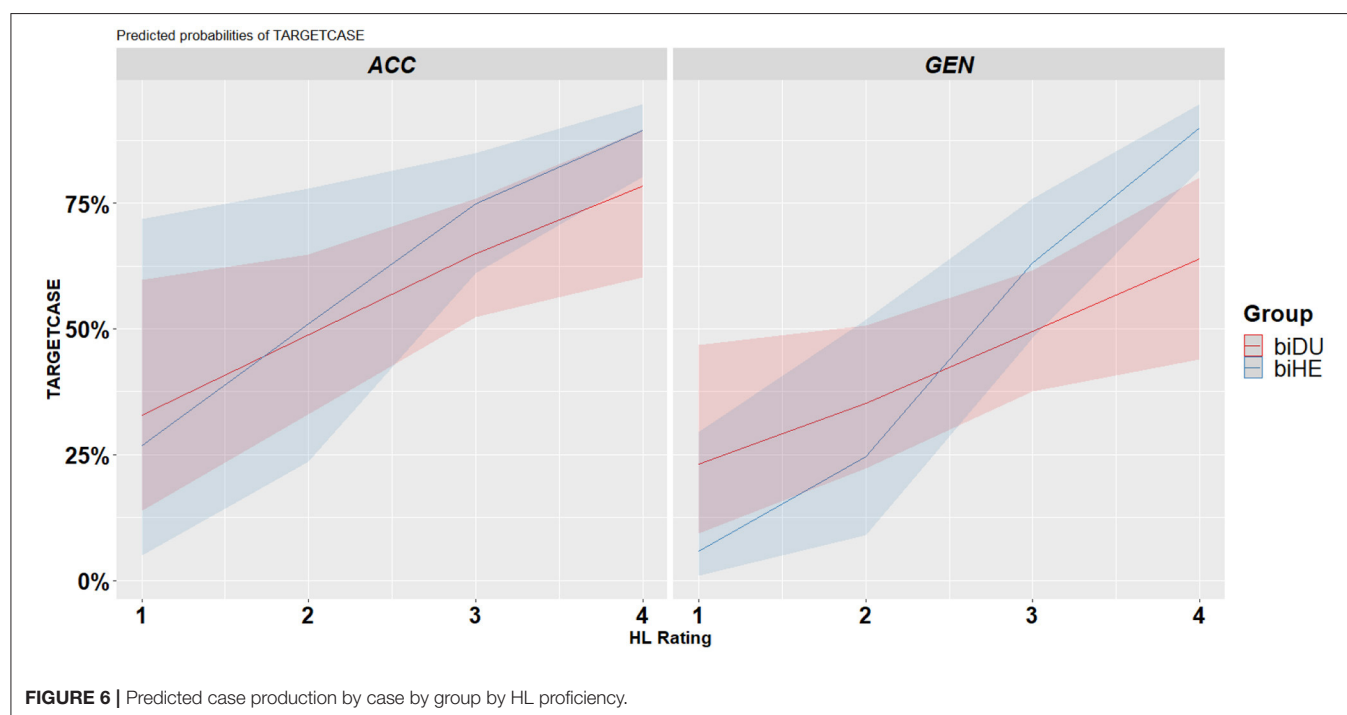


**FIGURE 6 |** Predicted case production by case by group by HL proficiency.

that the other indices of dominance (e.g., amount of language exposure and use, lexical proficiency, and fluency ratings by parents or teachers) did not detect the interaction between these dominance measures and cross-linguistic influence. In our study, the interaction between cross-linguistic influence and HL proficiency as indexed by parental ratings was observed. Our study showed that the growth is steeper across different proficiency levels when there is some overlap between the two languages. It should be noted that HL proficiency was closely related to the language spoken at home, AoO, and HL current input.

Previous attempts to model HL development, i.e., to distill the mechanisms influencing ultimate HL attainment, have arrived at different conclusions. Some scholars suggest that insufficient input is the key mechanism driving changes in HL grammar. Others believe that cross-linguistic influence is the sole mechanism shaping the ultimate HL attainment. The results of our study are in line with the proposals advocating for an integrative approach in modeling HL development. Cross-linguistic influences on HL are modulated by language-external factors (see Daskalaki et al., 2020; Van Dijk et al., 2021). Our findings show that being exposed to HL from birth does not guarantee the setting of parameters in line with monolingual grammar. The (re)-setting of parameters in HL is influenced by the presence/absence and realization of the parameters in the SL [as predicted by the *feature (re)-assembly hypothesis* (Lardiere, 2009)]. But we also showed that there are individual differences in the (re)-setting of the parameters, which are related to language-external factors. Putnam and Sánchez (2013) previously proposed that the (re)-setting of the feature parameters in HL is related to the feature activation in production and comprehension. The authors propose that HL speakers go through various stages in HL acquisition depending on the frequency of the feature activation. At Stage 1, HL speakers (re)-assemble some of their HL features in production under the influence and activation of SL features. At Stage 2, more HL features undergo re-assembly, including lexical items. At Stage 3, HL speakers show difficulties with HL features in production. Finally, at Stage 4, HL feature activation is impeded not only in production but also in comprehension. Our findings indirectly support this claim suggesting that the HL feature activation in production is very closely related to the characteristics of the language-external factors. HL speakers with more input/exposure have more opportunities to activate their HL features, as compared with those who have less input/exposure to HL and as a result, more input/exposure to SL. The decreased HL exposure and decreased HL proficiency affect the frequency of the activation of HL, which results in the re-assembly of features in some speakers under diminished HL input and subsequent lower language proficiency.

## CONCLUSIONS, APPLIED IMPLICATIONS, AND FUTURE RESEARCH

The results of the current study investigating morphosyntactic development in HL paint a complex picture. Our results show that HL acquisition is impeded under the influence of the

properties of SL which is evident in the lower performance of both bilingual groups as compared with the monolingual controls, who showed a near-ceiling performance on the production of target inflections in ACC and GEN contexts. More specifically, the acquisition of morphology is impeded when there are differences in the mapping of the functional features (as in the case with Russian-Hebrew bilingualism) and/or the absence of this feature marking (as in the case of Russian-Dutch bilinguals). The findings showed language-external factors as important factors modulating the potential negative effects of cross-linguistic influence. More specifically, the results show that AoO and HL proficiency play an important role in mitigating the possible effects of cross-linguistic influence. Finally, the study shows an interaction between SL properties and language-external factors: grammatical features which have more resemblance in the HL and the SL might be acquired faster under direct or indirect transfer from the SL. Thus, HL development is an intricate interplay between cross-linguistic influence and language-external factors.

The findings of the current study have important clinical and pedagogical implications. Firstly, on the clinical side, the study shows that the HL of bilingual children exhibits quantitative and qualitative divergences from the monolingual controls. Thus, clinicians need to be cautious when assessing and interpreting the assessment results of bilingual children in their HL (see Armon-Lotem et al., 2021). Secondly, the study shows a complex picture of HL development emphasizing the importance of HL input/exposure. Previous research showed that the parents of child HL speakers are generally interested in transmitting their HL to their children [see (Otwinowska et al., 2021)]—they see it as an integral part of maintaining a positive relationship with parents, grandparents, and extended families. Thus, efforts should be made to maintain HL input/exposure. Our findings confirmed that the HL “will not take care of itself” (Mieszkowska et al., 2017), as the HL is in danger of turning into a weaker language.

While our study makes a substantial contribution to the understanding of child HL development, it is not without limitations. First, our study focused only on the production of ACC and GEN cases, therefore, future studies should combine both production and comprehension tasks to deepen our understanding of the mechanisms of HL morphosyntactic acquisition. The investigation of production and comprehension of the same phenomenon in HL speakers would enable future studies to identify the developmental stage of the bilingual speaker. This would shed light on whether the problems are present only at the level of production, or if comprehension is also affected. Future research needs to focus more on the individual differences which shape HL development, including various linguistic, extra-linguistic, and cognitive measures. The inclusion of cognitive measures might help us understand how domain-general mechanisms (such as working memory and inhibition) modulate cross-linguistic competition. Finally, the current study was based on the participation of children with sparse case morphology in their SL. Future research should be extended to the bilingual-group comparisons of children speaking SLs that have rich case morphologies which are realized with case inflections (e.g., Finnish, Latvian, Estonian,

Ukrainian, Hungarian, Turkish), and SLs that mark case morphology differently from Russian (e.g., German, Greek). Studies comparing multiple bilingual groups will further deepen our understanding of the interplay between cross-linguistic influence and language-external factors.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Bar-Ilan University, The University of Amsterdam. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## REFERENCES

- Antonova Ünlü, E., and Li, W. (2018). The acquisition of the weaker language. *Ling. Approaches Bilingualism* 8, 637–663. doi: 10.1075/lab.16029.ant
- Armon-Lotem, S., and Meir, N. (2019). "The nature of exposure and input in early bilingualism," in *The Cambridge Handbook of Bilingualism*, eds A. De Houwer and L. Ortega (Cambridge: Cambridge University Press), 193–212.
- Armon-Lotem, S., Rose, K., and Altman, C. (2021). The development of English as a heritage language: the role of chronological age and age of onset of bilingualism. *First Lang.* 41, 67–89. doi: 10.1177/0142723720929810
- Babyonyshev, M. (1993). "The Acquisition of the Russian Case System," in *MIT Working Papers in Linguistics*, Vol. 19, 1–43.
- Bailyn, J. (2012). *The Syntax of Russian*. Cambridge: Cambridge University Press.
- Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: opportunities and challenges for linguistics. *Theor. Ling.* 39, 129–181. doi: 10.1515/tl-2013-0009
- Berman, R. A. (1978). *Modern Hebrew Structure*. Tel Aviv: University Publishing Projects.
- Blom, E., Cornips, L., and Schaffer, J. (eds.). (2017). *Cross-Linguistic Influence in Bilingualism: In Honor of Aafke Hulk*. Amsterdam: John Benjamins.
- Blom, E., Paradis, J., and Duncan, T. S. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular-s in child L2 English. *Lang. Learn.* 62, 965–994. doi: 10.1111/j.1467-9922.2012.00715.x
- Cejtlin, S. N. (2009). *Očerki po slovoobrazovaniju i formoobrazovaniju v detskoj reči*. Moskva: Znak.
- Chondrogianni, V., and Schwartz, R. G. (2020). Case marking and word order in Greek heritage children. *J. Child Lang.* 1–30. doi: 10.1017/S0305000919000849
- Cuza, A., and Pérez-Tattam, R. (2016). Grammatical gender selection and phrasal word order in child heritage Spanish: A feature re-assembly approach. *Bilingualism Language Cogn.* 19, 50–68. doi: 10.1017/S1366728914000893
- Cuza, A., Reyes, N., and Lustres, E. (2021). Copulas ser and estar production in child and adult heritage speakers of Spanish. *Lingua* 249:102978. doi: 10.1016/j.lingua.2020.102978
- Dąbrowska, E., and Szczerbiński, M. (2006). Polish children's productivity with case marking: the role of regularity, type frequency, and phonological diversity. *J. Child Lang.* 33, 559–597. doi: 10.1017/S0305000906007471
- Daskalaki, E., Chondrogianni, V., Blom, E., Argyri, F., and Paradis, J. (2019). Input effects across domains: the case of Greek subjects in child heritage language. *Second Lang. Res.* 35, 421–445. doi: 10.1177/0267658318787231
- Daskalaki, E., Elma, B. L. O. M., Chondrogianni, V., and Paradis, J. (2020). Effects of parental input quality in child heritage language acquisition. *J. Child Lang.* 47, 709–736. doi: 10.1017/S0305000919000850
- De Houwer, A., Miller, D., Bayram, F., Rothman, J., and Serratrice, L. (2018). "The role of language input environments for language outcomes and language acquisition in young bilingual children," in *Bilingual Cognition and Language: Bilingual Cognition and Language: The State of the Science Across its Subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins), 127–153.
- Gagarina, N. (2011). "Acquisition and loss of L1 in a Russian-German bilingual child: a case study," in *Monolingual and Bilingual Path to Language*, ed S. N. Cejtin (Moscow: Jazyki slavjanskij kul'tury), 137–163.
- Gagarina, N., and Klassert, A. (2018). Input dominance and development of home language in russian-german bilinguals. *Front. Commun.* 3:40. doi: 10.3389/fcomm.2018.00040
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingualism Lang. Cogn.* 12:213. doi: 10.1017/S1366728909004015
- Goebel-Mahrle, T., and Shin, N. L. (2020). A corpus study of child heritage speakers' Spanish gender agreement. *Int. J. Bilingualism* 24, 1088–1104. doi: 10.1177/1367006920935510
- Guiberson, M. M., Barrett, K. C., Jancosek, E. G., and Yoshinaga Itano, C. (2006). Language maintenance and loss in preschool-age children of Mexican immigrants: longitudinal study. *Commun. Disord. Quart.* 28, 4–17. doi: 10.1177/15257401060280010601
- Gvozdev, A. N. (1961). *Voprosy izučeniya detskoj reči*. Saint Petersburg: Detstvo-Press.
- Hulk, A., and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism* 3, 227–244. doi: 10.1017/S136672890000353
- Janssen, B. (2016). "The acquisition of gender and case in Polish and Russian: a study of monolingual and bilingual children," in *Pegasus Oost-Europese Studies*, Vol. 27 (Amsterdam: Uitgeverij Pegasus).
- Janssen, B., and Meir, N. (2019). Production, comprehension and repetition of accusative case by monolingual Russian and bilingual Russian-Dutch and Russian-Hebrew-speaking children. *Ling. Approaches Bilingualism* 9, 736–765. doi: 10.1075/lab.17021.jan
- Kaltsa, M., Prentza, A., and Tsimpli, I. M. (2020). Input and literacy effects in simultaneous and sequential bilinguals: the performance of Albanian-Greek-speaking children in sentence repetition. *Int. J. Bilingualism* 24, 159–183. doi: 10.1177/1367006918819867
- Kim, K., O'Grady, W., and Schwartz, B. D. (2018). Case in heritage Korean. *Linguistic Approaches to Bilingualism* 8, 252–282. doi: 10.1075/lab.16001.kim
- Kupisch, T., and Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Bilingualism* 22, 564–582. doi: 10.1177/1367006916654355
- Lardiere, D. (2009). Further thoughts on parameters and features in second language acquisition: a reply to peer comments on Lardiere's "Some thoughts on the contrastive analysis of features in second language acquisition in SLR 25". *Second Lang. Res.* 25, 409–422. doi: 10.1177/0267658309104579
- Lein, T., Rothweiler, M., and Hamann, C. (2017). "Factors affecting the performance in child heritage Portuguese in Germany," in *Proceedings of GALA*

## AUTHOR CONTRIBUTIONS

BJ carried out the collection of the data with monolingual participants and bilingual Russian-Dutch bilinguals. NM carried out data collection with bilingual Russian-Hebrew bilinguals and was responsible for data analysis and interpretation of the results. Both authors shared responsibility for drafting of the work, final approval of the version to be published, and the conception of the work.

## ACKNOWLEDGMENTS

We wish to thank the parents and the wonderful children who took part in this study. We also extend our gratitude to the two reviewers and the Editor of the paper for their most insightful and constructive comments on previous versions of this manuscript. We are also grateful to Maxwell Lichtenfeld and Sidney Gordon for their help in proofreading the manuscript.

- 2015, eds J. Choi, H. Demirdache, O. Lungu, and L. Voeltzel (Newcastle: Cambridge Scholars Publishing) 147–175.
- Meir, N., and Armon-Lotem, S. (2015). “Disentangling bilingualism from SLI in Heritage Russian: the impact of L2 properties and length of exposure to the L2,” in *Language Acquisition and Development: Proceedings of GALA 2013*, eds C. Hamann and E. Ruigendijk (Newcastle upon Tyne: Cambridge Scholars Publishing), 299–314.
- Meir, N., and Polinsky, M. (2021). Restructuring in heritage grammars: adjective-noun and numeral-noun expressions in Israeli Russian. *Linguist. Approach. Bilingual.* 11, 222–258. doi: 10.1075/lab.18069.mei
- Meir, N., Walters, J., and Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Ling. Approaches Bilingualism* 7, 514–553. doi: 10.1075/lab.15007.mei
- Mieszowska, K., Luniewska, M., Kolak, J., Kacprzak, A., Wodniecka, Z., and Haman, E. (2017). Home language will not take care of itself: vocabulary knowledge in trilingual children in the United Kingdom. *Front. Psychol.* 8:1358. doi: 10.3389/fpsyg.2017.01358
- Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139030502
- Montrul, S., and Sánchez-Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Lang. Acquis.* 20, 109–132. doi: 10.1080/10489223.2013.766741
- Ortega, L. (2020). The study of heritage language development from a bilingualism and social justice perspective. *Lang. Learn.* 70, 15–53. doi: 10.1111/lang.12347
- Otwinowska, A., Meir, N., Ringblom, N., Karpava, S., and La Morgia, F. (2021). Language and literacy transmission in heritage language: evidence from Russian-speaking families in Cyprus, Ireland, Israel and Sweden. *J. Multiling. Multicult. Dev.* 42, 357–382. doi: 10.1080/01434632.2019.1695807
- Polinsky, M. (2018). *Heritage Languages and Their Speakers*, Vol. 159. Cambridge: Cambridge University Press. doi: 10.1017/9781107252349
- Polinsky, M., and Scontras, G. (2019). Understanding heritage languages. *Bilingualism Lang. Cogn.* 22, 1–17. doi: 10.1017/S1366728919000245
- Polinsky, M., and Scontras, G. (2020). A roadmap for heritage language research. *Bilingualism Lang. Cogn.* 23, 50–55. doi: 10.1017/S1366728919000555
- Protassova, E., Mäki, M., and Rodina, N. (2017). Experimental investigation of Russian case acquisition among bilingual Finnish speaking children. *Acta Ling. Petropolitana* 13, 774–788.
- Protassova, E., and Voelkova, M. (2007). Diminutives in Russian at the early stages of acquisition. *Lang. Acquis. Lang. Disord.* 43:43. doi: 10.1075/lald.43.03pro
- Putnam, M. T., and Sánchez, L. (2013). What’s so incomplete about incomplete acquisition?: a prolegomenon to modeling heritage language grammars. *Ling. Approaches Bilingualism* 3, 478–508. doi: 10.1075/lab.3.4.04put
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/> (accessed October, 2021).
- Remennick, L. (2003). From Russian to Hebrew via HebRush: Intergenerational patterns of language use among former Soviet immigrants in Israel. *J. Multilingual Multicult. Dev.* 24, 431–453. doi: 10.1080/01434630308666509
- Ringblom, N. (2012). *The Acquisition of Russian in a Language Contact Situation: A Case Study of a Bilingual Child in Sweden*. Stockholm: Acta Universitatis Stockholmiensis.
- Rodina, Y., Kupisch, T., Meir, N., Mitrofanova, N., Urek, O., and Westergaard, M. (2020). Internal and external factors in heritage language acquisition: evidence from heritage Russian in Israel, Germany, Norway, Latvia and the United Kingdom. *Front. Educ.* 5:20. doi: 10.3389/educ.2020.00020
- Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: inflected infinitives in Heritage Brazilian Portuguese. *Int. J. Bilingualism* 11, 359–389. doi: 10.1177/13670069070110040201
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *Int. J. Bilingualism* 13, 155–163. doi: 10.1177/1367006909339814
- Schwartz, M., and Minkov, M. (2014). Russian case system acquisition among Russian-Hebrew speaking children. *J. Slavic Ling.* 22, 51–92. doi: 10.1353/jsl.2014.0005
- Schwartz, M., Minkov, M., Dieser, E., Protassova, E., Moin, V., and Polinsky, M. (2015). Acquisition of Russian gender agreement by monolingual and bilingual children. *Int. J. Bilingual.* 19, 726–752. doi: 10.1177/1367006914544989
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: determinants and mechanisms. *Ling. Approaches Bilingualism* 3, 3–25. doi: 10.1075/lab.3.1.01ser
- Serratrice, L. (2020). What counts as the baseline in child heritage language acquisition? *Bilingualism Lang. Cogn.* 23, 46–47. doi: 10.1017/S1366728919000518
- Sloussar, N. A., and Samoilova, M. V. (2015). Frequencies of different grammatical features and inflectional affixes in Russian nouns: a database.
- Sorace, A., Serratrice, L., Filiaci, F., and Baldo, M. (2009). Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children. *Lingua* 119, 460–477. doi: 10.1016/j.lingua.2008.09.008
- Treffers-Daller, J. (2019). What defines language dominance in bilinguals?. *Annu. Rev. Ling.* 5, 375–393. doi: 10.1146/annurev-linguistics-011817-045554
- Tsimpli, I. M. (2014). Early, late or very late?: timing acquisition and bilingualism. *Ling. Approaches Bilingualism* 4, 283–313. doi: 10.1075/lab.4.3.01tsi
- Tsinivits, D., and Unsworth, S. (2021). The impact of older siblings on the language environment and language development of bilingual toddlers. *Appl. Psycholinguist.* 42, 325–344. doi: 10.1017/S0142716420000570
- Turian, D., and Altenberg, E. P. (1991). “Compensatory strategies of child first language attrition,” in *First Language Attrition*, eds H. W. Seliger and R. M. Vago (Cambridge: Cambridge University Press), 207–226. doi: 10.1017/CBO9780511620720.014
- Unsworth, S. (2013). Current issues in multilingual first language acquisition. *Annu. Rev. Appl. Ling.* 33, 21–50. doi: 10.1017/S0267190513000044
- Unsworth, S. (2014). “Comparing the role of input in bilingual acquisition across domains,” in *Input and Experience in Bilingual Development*, eds T. Grüter and J. Paradis (Amsterdam: John Benjamins publishing House), 181–201. doi: 10.1075/tilar.13.10uns
- Unsworth, S. (2016). “Quantity and quality of language input in bilingual language development,” in *Lifespan Perspectives on Bilingualism*, eds E. Nicoladis and S. Montanari (Berlin: Mouton de Gruyter), 136–196.
- Unsworth, S. (2019). “Quantifying experience in heritage language development,” in *The Oxford Handbook of First Language Attrition*, eds M. Schmid and B. Köpcke (Oxford: Oxford University Press). doi: 10.1093/oxfordhb/9780198793595.013.34
- Van Dijk, C., Van Wonderen, E., Koutamanis, E., Kootstra, G. J., Dijkstra, T., and Unsworth, S. (2021). Cross-linguistic influence in simultaneous and early sequential bilingual children: a meta-analysis. *J. Child Lang.* 1–33. doi: 10.1017/S0305000921000337
- Vershina, E. A., and Eliseeva, M. B. (2007). “MakArturovskii oprosnik kak istochnik svedenii o rečevom razvitii rebenka,” in *Problemy ontolingvistiki-2007* (Sankt Peterburg: Materialy mežd. konf.), 72–76.
- Weinreich, U. (1968). *Languages in Contact. Findings and Problems*. Mouton: The Hague.
- Zdorenko, T., and Paradis, J. (2008). The acquisition of articles in child second language English: fluctuation, transfer or both?. *Second Lang. Res.* 24, 227–250. doi: 10.1177/0267658307086302

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Meir and Janssen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Modulation of Cross-Language Activation During Bilingual Auditory Word Recognition: Effects of Language Experience but Not Competing Background Noise

Melinda Fricke\*

Department of Linguistics, University of Pittsburgh, Pittsburgh, PA, United States

## OPEN ACCESS

### Edited by:

Tamar Degani,  
University of Haifa, Israel

### Reviewed by:

Alba Casado,  
Jagiellonian University, Poland  
Max Ryan Freeman,  
St. John's University, United States

### \*Correspondence:

Melinda Fricke  
melinda.fricke@pitt.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 February 2021

**Accepted:** 06 January 2022

**Published:** 24 February 2022

### Citation:

Fricke M (2022) Modulation of Cross-Language Activation During Bilingual Auditory Word Recognition: Effects of Language Experience but Not Competing Background Noise. *Front. Psychol.* 13:674157. doi: 10.3389/fpsyg.2022.674157

Previous research has shown that as the level of background noise increases, auditory word recognition performance drops off more rapidly for bilinguals than monolinguals. This disproportionate bilingual deficit has often been attributed to a presumed increase in cross-language activation in noise, although no studies have specifically tested for such an increase. We propose two distinct mechanisms by which background noise could cause an increase in cross-language activation: a phonetically based account and an executive function-based account. We explore the evidence for the phonetically based account by comparing cognate facilitation effects for three groups of native English listeners (monolinguals, late (L2) learners of Spanish, and heritage Spanish speakers) and four noise conditions (no noise, speech-shaped noise, English two-talker babble, and Spanish two-talker babble) during an auditory lexical decision task in English. By examining word recognition in the dominant language, the role of language control mechanisms is minimized, and by examining three different types of competing noise, the role of energetic vs. informational masking can be assessed. Contrary to predictions, we find no evidence that background noise modulates cross-language activation; cognate facilitation is constant across the four noise conditions. Instead, several indices of word recognition performance are found to correlate with aspects of linguistic experience: (1) The magnitude of the cognate facilitation effect is correlated with heritage listeners' self-ratings of Spanish proficiency; (2) Overall noise deficits are marginally larger for heritage listeners with lower English vocabulary scores; (3) Heritage listeners' Spanish self-ratings predict their magnitude of informational masking; (4) For all bilinguals, the degree of masking incurred in both English and Spanish two-talker babble is correlated with self-reported daily exposure to Spanish; and (5) The degree of masking incurred by Spanish babble is correlated with Spanish vocabulary knowledge. The results enrich our understanding of auditory word recognition in heritage speakers in particular and provide evidence that informational masking is most subject to modulation due to variation in linguistic experience. It remains to be seen whether cross-language activation is modulated by noise when the target language is the less dominant one.

**Keywords:** individual differences in language processing, speech perception in noise, auditory word recognition, heritage speakers, bilingualism, cognate effects

## INTRODUCTION

Non-native listeners are likely keenly aware that speech perception difficulties in adverse listening conditions appear amplified in a less-proficient language (see Garcia Lecumberri et al., 2010 and Scharenborg and van Os, 2019 for reviews). Indeed, findings have often indicated that second language listeners suffer disproportionately in noise as compared to their native language counterparts, especially in tasks involving whole-word perception (Black and Hast, 1962; Cooke et al., 2008; Morini and Newman, 2020). There are many reasons for this: non-native listeners are likely to have less robust phonetic (Hazan and Simpson, 2000; Cutler et al., 2004, 2008), phonological (Weber and Cutler, 2004; Broersma and Cutler, 2011), and lexical (Gollan et al., 2008; Gollan et al., 2011a; Shook et al., 2015) representations than natives, and as a result, they are more likely to experience bottlenecks in linguistic processing (Krizman et al., 2017), all of which may also adversely impact their ability to take advantage of contextual (Bradlow and Alexander, 2007; Skoe and Karayanidi, 2019) and semantic (Golestani et al., 2009) information so as to offset processing difficulties at other levels of representation.

Bilingual listeners also have to contend with cross-language activation from the non-target language. Decades of research on bilingual word recognition, both visual and auditory, have demonstrated that cross-language activation is pervasive in bilingual processing (e.g., Caramazza and Brones, 1979; Spivey and Marian, 1999; Van Hell and Dijkstra, 2002; Lagrou et al., 2011). Noisy listening conditions would only seem to exacerbate this issue, and indeed, many researchers have suggested exactly this (Rogers et al., 2006; Krizman et al., 2017; Morini and Newman, 2020). To our knowledge, however, no study has directly supported this hypothesis; other than speech perception in noise (“SPIN”) being more difficult in a non-native language, we are aware of no direct evidence that cross-language activation processes are altered by the presence of background noise.

This is the primary research question of the current study: *are* cross-language activation processes modulated by the presence of background noise, and if so, under what circumstances and for which listeners?

## BACKGROUND

### Bilingual Speech Perception in Noise

Following recent calls to treat language experience-related predictors as continuous, rather than categorical, variables (Luk and Bialystok, 2013; Birdsong, 2018), we use the term “bilingual” broadly to refer to any individuals with knowledge of more than one language. We acknowledge, however, that this usage is distinct from much of the literature cited here.

The vast majority of studies examining bilingual SPIN has focused on perceptual deficits in a non-natively acquired language. Listeners who acquired the target language later in life are typically more strongly affected by adverse listening conditions as compared to listeners who acquired the language earlier, and this generalization holds for comparisons of natives vs. non-natives (Garcia Lecumberri and Cooke, 2006;

Scharenborg et al., 2018) as well as earlier vs. later L2 learners (Mayo et al., 1997; Meador et al., 2000). Several studies have specifically implicated age of acquisition (“AoA”) as an explanatory variable (Mayo et al., 1997; MacKay et al., 2001), but given nearly unavoidable confounds between AoA and other potentially explanatory variables (dominance, proficiency, length of exposure, and context of acquisition), more work is still needed to understand their independent contributions, an issue to which we return below.

SPIN deficits also vary depending on the level of linguistic processing. Studies focusing on lower level phonetic perception, such as consonant and vowel identification, have most often found that the deficit for non-native as compared to native listeners remains constant even as the amount of noise increases (Takata and Nábelek, 1990; Hazan and Simpson, 2000; Cutler et al., 2004). Studies investigating higher level processing, on the other hand, have observed disproportionate deficits for non-natives. Tasks involving word recognition, whether in a sentence context (Mayo et al., 1997; Bradlow and Alexander, 2007) or not (Tabri et al., 2011; Scharenborg et al., 2018; Morini and Newman, 2020), have more often found that in increasing levels of noise, non-native performance drops off at a faster rate relative to natives. This suggests that bilingual *lexical* processing may be particularly disrupted in noise and/or that focusing on word recognition in noise may provide a window into a critical nexus of processing demands.

This latter point also intersects with an important distinction made in the literature on SPIN, that of energetic vs. informational masking (Cooke et al., 2008). Energetic masking refers to the idea that interfering sounds can render the target speech inaudible due to overlapping time and frequency characteristics of the sounds (e.g., being unable to hear a word because of a train rushing by), while informational masking refers to the decrement in recognition performance that occurs due to the informational content of non-target sounds (e.g., the ability of nearby speech to capture the listener’s attention). As such, informational masking is a catch-all term that encompasses a variety of cognitive-linguistic phenomena that do not necessarily form a unitary construct.

In an effort to understand the components of informational masking, studies have varied both the informational content of the masking sounds and the linguistic background of the listeners. Such work has demonstrated that listeners’ facility in deriving informational content from competing sounds plays a role in the degree of informational masking they experience. While it is known that a competing talker provides the equivalent of 6–8 dB *less* masking as compared to constant (“stationary”) noise at the same level (Festen and Plomp, 1990), the reason for this is that in the presence of competing speech, the target signal can still be “glimpsed” (Peters et al., 1998; Cooke, 2006), i.e., the time-frequency characteristics of speech are such that only bits and pieces of the target speech will be covered up by the competing speech at any moment in time. When the benefit afforded by glimpsing is taken into account, competing speech is generally *more* disruptive than non-speech noise, but also *variably* more disruptive. The reason is that cognitive factors, such as the ability to maintain attention on the target

speech and suppress the activation of non-target representations, become a crucial determinant of the degree of informational masking experienced by the listener. Consequently, the maximum degree of informational masking has been reported to occur with two competing talkers (Freyman et al., 2004), because as the number of competing talkers increases, listeners' capacity to derive informational content from the masker becomes swamped.

In the same vein, proficiency in the language of the masker also plays a role. Work by Van Engen and Bradlow (2007) and Van Engen (2010) demonstrated that for both native and non-native listeners, the ability to understand competing speech is associated with reduced word recognition accuracy. And while these studies compared listeners with either native or no proficiency in the masking language (see also Cooke, 2006), some studies have also found differential susceptibility to disruption from competing speech as a function of smaller gradations of proficiency in the masking language (Imai et al., 2005; Kilman et al., 2014).

Importantly, however, most studies examining the role of language experience in SPIN have involved group-level comparisons, sometimes with relatively little detail given about the participants themselves, and often relying on self-ratings of language ability rather than objective measures of language knowledge (but see Ezzatian et al., 2010; Van Engen, 2010; Warzybok et al., 2015; Scharenborg et al., 2018). As a result, it is as yet largely unknown which aspects of differential performance between listener groups should be attributed to differences in age of acquisition *per se* versus dominance, proficiency (and if so, which aspects), or domain-general cognitive abilities. In short, the *mechanisms* relating language experience to SPIN are not yet clear, and this will remain the case until the multitude of factors related to language experience can be disentangled.

## The Special Case of Heritage Listeners

In line with our broad usage of the term “bilingual,” we adopt the similarly broad definition of heritage bilingualism put forth by Rothman (2009), who proposes that a heritage language is “a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society,” and a heritage speaker is someone who “has some command of the heritage language acquired naturalistically (e.g., Valdés, 1995, 2000).” This usage encompasses both simultaneous and sequential child bilingualism, and it acknowledges the fact that the distinction between these categories is in many cases blurred, depending on how many members of the family and immediate social network are fluent in the minority and/or majority languages. What heritage speakers of all profiles have in common is that they typically become more dominant in the societal majority language after beginning schooling. Heritage bilinguals' language processing behavior is therefore likely to be similar in some ways to that of monolinguals (due to their early AoA and implicit acquisition) and in other ways to that of L2 late learners (due to their varying degrees of experience and ultimate attainment), but the details of processing behavior

in this population remain poorly understood. Moreover, while these observations have been made in the literature in reference to knowledge of the heritage language itself (Montrul et al., 2008; Bolger and Zapata, 2011), especially in the context of SPIN, it bears some discussion that they could equally apply to the later-acquired majority language; we return to this point below.

In the United States, heritage speakers of Spanish generally acquire Spanish in the home from birth, and later become more dominant in English due to a relative lack of societal and community support for Spanish language use (Rothman, 2009). They may begin learning English at birth, simultaneously with Spanish, or slightly later, in young childhood, as their exposure to mainstream, English-dominant culture increases. Given the language context of the United States, the vast majority will ultimately participate in the English-speaking school system and workforce, and as a result, the quantity and quality of English exposure over the life span will be relatively comparable across individuals, with the main differences in English exposure relating to the nature of English input received before the onset of schooling around age four to five. The profile of Spanish experience across the life span, by contrast, may differ more dramatically depending on individual circumstances.

Investigating language processing behavior in US heritage speakers of Spanish therefore affords a unique opportunity to observe how the psycholinguistic processing of a dominant language can be modulated by linguistic experience in a non-dominant language, when differences in AoA for the two are minimized. This question is of great interest in the context of recent perspectives on bilingualism that place the role of linguistic experience and plasticity across the life span front and center for advancing our understanding how the cognitive system accommodates the presence of multiple languages (Baum and Titone, 2014; Kroll et al., 2014). From this standpoint, it is somewhat surprising that relatively little is known regarding language processing in heritage Spanish speakers. While some recent studies have investigated Spanish sentence processing in this population (Jegerski, 2018; Jegerski and Sekerina, 2020), and one study compared Spanish heritage speakers' English sentence processing to that of English monolinguals (Bice and Kroll, 2021), most investigations of heritage speakers' linguistic abilities have focused on offline knowledge of Spanish grammatical structures, using tasks, such as grammaticality judgments or structure elicitation tasks (e.g., Montrul, 2009; Montrul and Bowles, 2009). As a result, little is known about how factors such as dominance and proficiency influence basic online language processing in this population, in either Spanish or English.

One partial exception to this statement, however, is heritage Spanish speakers' English auditory word recognition in adverse listening conditions. Perhaps surprisingly, given their lifetime of experience and consequent dominance in English, heritage Spanish listeners have repeatedly shown deficits in English SPIN relative to monolinguals. Two widely cited early studies found that despite acquiring English before age six and demonstrating monolingual-like performance (at ceiling) in the clear, heritage Spanish listeners'

English word recognition abilities in noise lagged behind that of their monolingual English counterparts. Mayo et al. (1997) found that heritage listeners required more favorable signal-to-noise ratios (SNRs) and derived less benefit from context as compared to monolinguals, and Rogers et al. (2006) found lower word shadowing accuracy at three SNRs and in two different adverse listening conditions. See also Tabri et al. (2011) for similar results in a somewhat comparable group of Arabic-English early bilinguals. In both of these studies, the patterns seemed to hold whether participants had acquired English simultaneously with Spanish since birth or had begun learning English slightly later, but both studies suffered from small sample sizes, with just 12 heritage bilinguals in each. Moreover, neither study arguably included any objective, independent measures of linguistic proficiency, making it difficult to identify the source of the observed deficits. Rogers et al. did report, though, that their heritage Spanish and monolingual English groups were matched on English accentedness, suggesting that English phonetic knowledge was at least somewhat comparable across groups.

More recently, two studies including more participants and more measures of linguistic performance have further explored heritage listeners' English word recognition in noise. Krizman et al. (2017) tested 25 adolescent heritage bilinguals on a battery of perceptual tasks. They found that heritage bilinguals performed worse than monolinguals for sentence perception in noise, equivalently to monolinguals for single word perception in noise, and better than monolinguals for pure tone detection in noise. Morini and Newman (2020) tested 32 heritage bilinguals' word recognition and word learning abilities in noise and found bilingual deficits (relative to monolinguals) only in the recognition task. Taken together, these studies support the idea that processes involved in retrieving linguistic representations from memory may be a significant source of difficulty for bilinguals charged with processing speech in noise, but they leave open the question of whether factors known to impact the retrieval process, such as the strength of cross-language activation and/or proficiency in the non-dominant language, may interact with the presence of noise.

Intriguingly, one small study found a relationship between L2 proficiency and SPIN performance in L1 for late L2 learners: von Hapsburg and Bahng (2009) reported that among native Korean late learners of English, greater L2 English proficiency was associated with worse L1 Korean word recognition in noise. This is particularly striking in the context of the plasticity-oriented perspectives highlighted above (Baum and Titone, 2014; Kroll et al., 2014). The process of acquiring and strengthening L2 representations and processing routines requires the learner to integrate new information into an already established system, a process that entails adaptation of the L1. As such, the development of proficiency in the L2 is ultimately a question of plasticity of the language processing system, i.e., the flexibility of the cognitive architecture that supports both the native language and any subsequently learned languages. While it is already widely accepted that the L1 should have a strong influence on the L2, perhaps especially in contexts of increased processing demands, such as adverse listening conditions, what this perspective underscores is that successful

L2 acquisition may in some cases be associated with less optimal L1 performance.

To investigate these issues in more detail, the present study compares English word recognition in heritage Spanish bilinguals to that of both monolingual English and native English-late L2 Spanish bilinguals. Before introducing the study, however, we first provide an overview of what is known regarding cross-language activation in bilingual auditory word recognition.

## Cross-Language Activation in Bilingual Auditory Word Recognition

While the majority of evidence for cross-language activation in bilingual language processing comes from studies of visual word recognition (see Dijkstra and van Heuven, 2018, for a recent review), there is also considerable evidence that bilinguals experience cross-language activation in the auditory modality. Many studies in this vein have employed the visual world paradigm, often finding that bilingual listeners are more likely to look at an interlingual distractor picture (e.g., a duck—Spanish *pato*—when the target word is English “pot”) as compared to unrelated distractor pictures (Spivey and Marian, 1999; Marian and Spivey, 2003). Such findings indicate that representations in the non-target language are activated in the course of recognizing words in the target language. While cross-language activation appears robust when the non-target language is the dominant one (Blumenfeld and Marian, 2007; Chambers and Cooke, 2009), some studies have reported non-target activation of the non-dominant language as well, and it is likely that the presence of cross-language activation when the non-target language is *non-dominant* depends on factors such as AoA/proficiency (Canseco-Gonzalez et al., 2010) and the language immersion context (Spivey and Marian, 1999, whose participants had been immersed in the L2 for an average of 4 years).

There is some evidence for cross-language activation during auditory word recognition from non-visual world paradigms as well, but the data are actually rather sparse concerning auditory recognition of the type of “between language” words that share extensive cross-language overlap and that have often been employed in the literature on visual word recognition. The logic in these studies is that to the extent that representations in the non-target language become active in the course of target word recognition, words that overlap in form across languages should show differential processing as compared to control items. This prediction has been borne out in the auditory modality in several studies that have found differential recognition of interlingual homophones (Schulpen et al., 2003; Lagrou et al., 2011) and cognates (Woutersen et al., 1995; Blumenfeld and Marian, 2007; Guediche et al., 2020) as compared to control words, but on the whole, more data are needed in order to understand the relationships among cross-language overlap, language-specific phonetic cues, and cross-language activation patterns in the auditory modality.

## Mechanisms by Which Cross-Language Activation Could Increase in Noise

In the auditory modality, language-specific phonetic cues could help bilinguals restrict activation to representations

in the target language. Even words that share the same coarse-grained phonemic units across languages will generally be pronounced in such a way as to make the intended language clear, i.e., cognates and interlingual homophones will be realized with different “accents” depending on the language being spoken. The empirical record is quite mixed, however, with a handful of studies reporting that participants could take advantage of such language-specific phonetic cues (Schulpen et al., 2003; Ju and Luce, 2004; Fricke et al., 2016), and others finding they could not (Lagrou et al., 2011; McDonald and Kaushanskaya, 2020). Critically for the present study, even if some participants are capable of exploiting language-specific cues when listening conditions are favorable, the ability to do so should be greatly reduced in the presence of competing noise as a result of energetic and/or informational masking (Mattys et al., 2014). This should in turn lead to increased activation of words in the non-target language. We refer to this hypothesis as a *phonetically based account* of increased cross-language activation in the presence of background noise: to the extent that noise makes bilingual listeners unable to exploit phonetic cues to language membership, competing lexical representations in the non-target language should be more active in noise as compared to in the clear.

An alternative, though not mutually exclusive, possibility is that the necessity of directing cognitive resources toward the tasks of isolating and tracking the target speech stream could reduce the resources available for language control processes. Noisy listening conditions are understood to qualitatively alter the dynamics of lexical competition in the native language (McQueen and Huettig, 2012; Brouwer and Bradlow, 2016; Scharenborg et al., 2018) and recent perspectives on SPIN place the role of cognitive load front and center (see Peelle, 2018, and Pichora-Fuller et al., 2016 for reviews). During bilingual word recognition, inhibitory control in particular is known to play a role in resolving cross-language competition (Blumenfeld and Marian, 2013; Mercier et al., 2014; Chen et al., 2017), and it remains an open question to what extent other aspects of executive function may be involved as well (Kroll and Bialystok, 2013; Antoniou, 2019). An alternative to the phonetically based account is therefore an *executive function-based account*: to the extent that comprehending speech in noise taxes the cognitive system, fewer cognitive resources may be available for managing activation of the non-target language, resulting in greater cross-language activation in noise.

Importantly, the presence or absence of increased cross-language activation in noise will not enable us to distinguish between these two accounts. However, since language control processes are more likely to be engaged when listening in the less dominant language (i.e., when the non-target language is more dominant; Mercier et al., 2014; see also Green, 1998 and Misra et al., 2012), the present study provides a stronger test of the phonetically based account. By examining recognition of the dominant language by proficient speakers of a non-dominant language, the current study tests the effects of noise on cross-language activation while minimizing the role of language control processes.

## The Present Study

The present study investigates English auditory word recognition in three groups of listeners, all self-identified native English speakers, in the presence of three types of competing noise: speech-shaped noise, English two-talker babble, and Spanish two-talker babble. The inclusion of three distinct populations of native English listeners allows us to explore the questions of how proficiency and context of acquisition of the non-dominant language impact cross-language activation processes in noise, and the inclusion of three distinct types of noise allows us to examine the impact of the content of the noise itself.

The strength of cross-language activation was operationalized by measuring the extent of any cognate facilitation effects. We predicted that an increase in non-target language activation would boost recognition accuracy and speed for cognates relative to control words, such that any deleterious effects of noise on word recognition would be less severe for cognates relative to controls. Thus, while cognates should still be recognized with more difficulty in noise as compared to in the clear, the decrement in recognition performance may be less for cognates relative to control words.

The focus on native language processing in this study has several motivations. For one, the question of variation in native language processing as a function of language experience is compelling from a plasticity-oriented viewpoint. The present listeners all self-identify as native English speakers and consider English their more dominant language. The measures on which they differ most dramatically concern their experience with Spanish, allowing for an exploration of how aspects of this experience might impact processing in the more dominant language. The inclusion of two relatively large and qualitatively distinct bilingual participant groups also enables both group-based and individual differences-based analyses, helping to clarify whether differences across groups are easily captured by existing metrics (e.g., if the roles of AoA, lexical proficiency, etc. are constant across groups), or to the extent that they are not, suggesting avenues for future research. Finally, the present study has implications for bilinguals with native or native-like proficiency in more than one language. Heritage speakers of Spanish make up a significant proportion of the US population; 13.4% of the population aged 5 years or older is currently estimated to speak Spanish at home (U.S. Census Bureau, 2019). Understanding the extent to which language processing behavior in this population is comparable to that of monolingual native English speakers is not only essential to ensuring best practices in clinical and policy decisions, it also promises to enrich our understanding of the basic mechanisms governing language acquisition, speech perception, and cognitive adaptability.

## MATERIALS AND METHODS

### Participants

Participant recruitment took place *via* Prolific.<sup>1</sup> The study advertisement was shown only to Prolific users who met the

<sup>1</sup>[www.prolific.co](http://www.prolific.co)

following screening criteria: age 18–35, born in the United States, currently living in the United States, and reporting English as (one of) their native language(s). To identify potential monolingual participants, the following additional criteria were also applied: self-reported fluency in English only and raised in a monolingual environment (“I was raised with my native language only.”). To identify potential L2 Spanish speakers, we screened for participants who reported being fluent in one or more languages in addition to English, were raised in a monolingual environment, and also reported being fluent in Spanish. Finally, to identify potential heritage speakers of Spanish, the following additional screening criteria were applied: US citizen, fluent speaker of one language in addition to English, raised with two or more languages spoken in the home, and fluent in Spanish.

Potential participants identified by these Prolific-internal screening procedures first participated in a study-specific screening session in which they completed the English LexTALE (Lemhöfer and Broersma, 2012), Spanish LexTALE (Izura et al., 2014), and the LEAP-Q (Marian et al., 2007). The LexTALE is a brief (3–4 min) lexical decision task, independently created and normed for each language, providing an objective measure of vocabulary knowledge. The measure used in all LexTALE analyses was the “Average Percent Correct” (Lemhöfer and Broersma, 2012), the grand mean of the average percent correct on word trials and nonword trials. Monolinguals were invited to participate in the main study if they reported Spanish comprehension ability of 3 or less on a scale of 0 (no knowledge) to 10 (like a native speaker), while L2 and heritage participants were initially invited to participate if they reported a 7 or higher. Following an initial period of recruitment (20–25 participants per group), only bilingual participants who achieved at least 60% on the Spanish LexTALE were invited to participate, in order to facilitate regression-based analyses involving this variable, with a final target group size of around 30 per group.

A total of 101 participants completed the main experiment. All reported normal hearing and no history of speech or language disorder. Ten participants were excluded due to low effort responses in the experimental task (defined as either no response or response times faster than 300 ms for more than 25% of trials), one was excluded due to ambiguous responses on the LEAP-Q, and one was excluded for being an early English-Mandarin bilingual, leaving a total of 89 participants’ data for analysis. The LEAP-Q and LexTALE data for the final sample of participants are summarized in **Table 1**. Several heritage participants had missing data for the question concerning how many years they had spent in an English- (five participants) or Spanish-speaking (three participants) household. The latter three participants were excluded from the individual differences analyses but were included in the group analyses because they listed Spanish as a language spoken in their home growing up.

The three participant groups were compared using one-way ANOVAs. These returned differences in age [ $F(2,86)=4.7$ ,  $p=0.01$ ], English age of acquisition [ $F(2,86)=10.9$ ,  $p<0.001$ ], daily exposure to English [ $F(2,86)=19.9$ ,  $p<0.001$ ], number of years in an English-speaking household [ $F(2,81)=10.0$ ,  $p<0.001$ ], and all of the measures related to Spanish experience (all  $F_s>20.0$ , all  $p_s<0.001$ ). There was a marginal difference

in English LexTALE scores [ $F(2,86)=3.1$ ,  $p=0.05$ ]. Two-tailed Welch-corrected  $t$ -tests with a Bonferroni-adjusted  $\alpha$  of 0.017 were used to determine which pairwise group comparisons were significant. For age, the Heritage group was slightly younger than both the Monolingual [ $t(55.0)=-2.6$ ,  $p=0.01$ ] and L2 group [ $t(58.5)=-2.8$ ,  $p=0.009$ ]. For the English measures, the Monolingual and L2 group differed only in their daily exposure to English [ $t(48.1)=-2.7$ ,  $p=0.009$ ]. The Heritage group reported a later English AoA than both the Monolingual [ $t(36.2)=3.5$ ,  $p=0.002$ ] and the L2 [ $t(32.1)=3.6$ ,  $p<0.001$ ] groups, as well as fewer years spent in an English-speaking household and less daily English exposure than both the Monolingual [ $t(35.3)=-3.5$ ,  $p=0.001$ ;  $t(50.5)=-6.9$ ,  $p<0.001$ ] and L2 groups [ $t(34.8)=-3.1$ ,  $p=0.004$ ;  $t(58.8)=-3.3$ ,  $p=0.001$ ]. The Heritage group’s English LexTALE scores were marginally lower than that of the Monolinguals [ $t(50.9)=-2.3$ ,  $p=0.02$ ] and the L2 group [ $t(57.4)=-1.8$ ,  $p=0.08$ ]. For all of the Spanish measures, both bilingual groups differed significantly from the Monolinguals (all  $|t|s>2.7$ , all  $p_s<0.01$ ), with the exception of Spanish AoA, where the L2 group did not differ from the subset of 10 Monolinguals who had studied Spanish [ $t(27.6)=-0.38$ ,  $p=0.71$ ], but the Heritage group did [ $t(10.3)=12.3$ ,  $p<0.001$ ]. The two bilingual groups differed from one another for all Spanish measures (all  $|t|s>2.5$ , all  $p_s<0.016$ ) with the exception of self-rated reading [ $t(59.0)=0.4$ ,  $p=0.69$ ] and Spanish LexTALE scores; the latter comparison was marginal [ $t(40.6)=2.0$ ,  $p=0.06$ ]. No other differences approached significance.

The monolingual group was on the whole quite monolingual. Just 10 of 28 reported having studied Spanish in school, with an average AoA of 12.6 ( $SD=3.0$ ) and an average composite Spanish self-rating of 0.4 ( $SD=0.5$ ), while 10 of 28 reported experience with a language other than Spanish, with an average AoA of 15.6 ( $SD=4.9$ ) and an average composite self-rating of 1.9 ( $SD=1.7$ ).

## Procedure

### Experimental Session

All experimental procedures were conducted using Gorilla Experiment Builder<sup>2</sup> (Anwyl-Irvine et al., 2020), and participants completed them *via* the Internet in a location of their choosing. Participation was restricted to users of desktop computers rather than mobile devices to maximize the probability that participants would be seated, in a location with minimal distractions.

The experiment reported here was the second experimental task of the session. After giving informed consent, participants completed a six-trial headphone check (Woods et al., 2017). The first experimental task in the session was a word transcription task, followed by the lexical decision task reported here, followed by a phonetic perception task, and followed finally by the AX variant of the Continuous Performance Task (Braver et al., 2001; Morales et al., 2013). The full session took approximately 90 min, and participants were compensated \$15 for their time with a \$1 bonus for successfully completing all tasks in the session.

<sup>2</sup>www.gorilla.sc

**TABLE 1 |** Summary of participant characteristics (means and SDs).

|                                            | Monolingual English |             | L2 Spanish         |                    | Heritage Spanish   |                    |
|--------------------------------------------|---------------------|-------------|--------------------|--------------------|--------------------|--------------------|
| N (N female)                               | 28 (12 F)           |             | 30 (18 F)          |                    | 31 (12 F)          |                    |
| Age                                        | 28.0 (5.1)          |             | <b>27.9 (5.0)</b>  |                    | <b>24.6 (4.7)</b>  |                    |
|                                            | Eng                 | Spa         | Eng                | Spa                | Eng                | Spa                |
| Self-rated comprehension                   | 10.0 (0.2)          | 0.5 (0.6)   | 10.0 (0)           | <b>6.9 (2.0)</b>   | 10.0 (0)           | <b>9.0 (1.4)</b>   |
| Self-rated speaking                        | 10.0 (0)            | 0.2 (0.4)   | 10.0 (0.2)         | <b>6.4 (1.8)</b>   | 9.9 (0.4)          | <b>7.5 (1.6)</b>   |
| Self-rated reading                         | 10.0 (0.2)          | 0.3 (0.6)   | 10.0 (0)           | 7.1 (2.1)          | 10.0 (0)           | 7.3 (2.2)          |
| Self-ratings composite                     | 10.0 (0.1)          | 0.4 (0.5)   | 10.0 (0.1)         | <b>6.8 (1.7)</b>   | 10.0 (0.1)         | <b>7.9 (1.4)</b>   |
| Age of acquisition                         | 0.2 (0.7)           | 12.6* (3.0) | <b>0.1 (0.4)</b>   | <b>13.1 (5.1)</b>  | <b>1.5 (2.2)</b>   | <b>0.7 (1.4)</b>   |
| # yrs. in household where this is spoken** | 28.0 (5.1)          | 0 (0)       | <b>27.0 (5.1)</b>  | <b>1.1 (2.1)</b>   | <b>19.8 (10.7)</b> | <b>21.1 (6.4)</b>  |
| Percent daily exposure***                  | 93.5 (9.6)          | 3.8 (7.4)   | <b>84.2 (15.9)</b> | <b>12.9 (13.6)</b> | <b>70.5 (15.6)</b> | <b>24.6 (13.6)</b> |
| LexTALE                                    | 95.0 (4.7)          | 47.6 (5.9)  | 94.4 (6.2)         | 59.1 (6.6)         | 91.2 (7.6)         | 65.1 (15.7)        |

Bolded cells are those for which means for the two bilingual groups differed (adjusted  $\alpha=0.017$ ). See text for more detailed comparisons.

\*Average for participants who entered a response ( $n=10$ ). \*\*See note in the text regarding missing data. \*\*\*Some participants did not give percentages that summed to 100; these were rescaled to add up to 100.

## Design

The lexical decision task comprised a total of 240 trials, divided into four blocks of 60 trials each, with six practice trials at the beginning of each block and an opportunity to take a short break in between blocks. Each block consisted of half real English words and half nonwords derived from English words, and the 30 word trials in each block consisted of half English-Spanish cognates and half non-cognates (see “Materials”). Each of the four blocks constituted a different noise condition: the first block was completed in the clear (i.e., no background noise), followed by a block with speech-shaped noise (SSN), a block with English two-talker babble (E2TB), and finally a block with Spanish two-talker babble (S2TB). Block order was fixed to keep any ordering effects constant across participants rather than further complicate the design. The E2TB block was ordered before the S2TB block to prevent any carryover effects of cross-language activation from one block to the next (Misra et al., 2012). The full set of stimuli was divided into four sets, with the cognates and non-cognates within each set matched as closely as possible (see “Materials”). Four different versions of the experiment were created, rotating the item sets through the different noise conditions, to ensure that any cognate effects were not dependent on the specific items in a given condition. Trial order was fully randomized within each block.

## Trial Procedure

On each trial, participants heard a single word or nonword embedded in the carrier phrase, “Now I’ll say...” and were asked to “determine whether the last item in the sentence is a real English word or a made-up word.” Participants were asked to use their left hand to press “1” on their computer keyboard to respond “real word” or their right hand to press “0” to respond “not a word.” The response options were displayed on the screen throughout the task. For trials in the noise blocks, the noise started 500ms before the onset of the carrier phrase and continued until 500ms after the offset of the target word. Participants had up to 3,000ms following the onset of

the target word to make a response, at which point the words “Time’s up! Try to respond faster!” were displayed on the screen.

## Materials

The full list of experimental stimuli is available at <https://osf.io/t9prb>. The stimuli were compiled by two research assistants with language backgrounds equivalent to the bilingual participants in the experiment (both native English, proficient in Spanish, one a late L2 learner, and one a heritage speaker of Spanish), under the supervision of the author. All stimuli were two syllables long, with stress overwhelmingly on the first syllable. The nonwords were loosely based on the real words and were created by altering two or more phonemes of each word stimulus so as to obscure the relationship between the word and its derived nonword. For example, the nonword *reckle* was derived from the target word *metal*. None of the target words appeared within the same block as their derived nonword.

Cognate and noncognate stimuli were matched on the following attributes (see Table 2; all statistics obtained from the CLEARPOND lexical database; Marian et al., 2012): log-transformed word frequency (from the SUBTLEX-US corpus; Brysbaert and New, 2009), length in phonemes, number of English and Spanish phonological neighbors (most words had no Spanish phonological neighbors), and the mean positional frequency and biphone frequency of all phonemes/biphones in the word. Two-tailed Welch-corrected *t*-tests comparing cognates to noncognates for each of the four stimulus subsets separately, and the full stimulus set combined, confirmed that stimuli did not differ along any of these dimensions (all  $|t|s < 1.7$ , all  $ps > 0.10$ ). However, two-tailed continuity-corrected Wilcoxon tests examining word durations indicated that cognate stimuli were overall longer than noncognates ( $W=2,354$ ,  $p=0.004$ ); this difference was also significant for Set 1 ( $W=164$ ,  $p=0.03$ ) and marginal for Set 2 ( $W=160$ ,  $p=0.05$ ).

The stimuli were recorded by a native speaker of Central American Spanish whose English proficiency and accent were subjectively native-like. This speaker was chosen for her ability to record native-sounding stimuli for both the English and Spanish

**TABLE 2 |** Summary of stimulus characteristics (means and *SDs*, real words only).

|                    | Set 1                                        |                      | Set 2                                       |                      | Set 3                                        |                      | Set 4                                        |                      | All                                                       |                      |
|--------------------|----------------------------------------------|----------------------|---------------------------------------------|----------------------|----------------------------------------------|----------------------|----------------------------------------------|----------------------|-----------------------------------------------------------|----------------------|
|                    | Cog ( <i>n</i> = 15)                         | Non ( <i>n</i> = 15) | Cog ( <i>n</i> = 15)                        | Non ( <i>n</i> = 15) | Cog ( <i>n</i> = 15)                         | Non ( <i>n</i> = 15) | Cog ( <i>n</i> = 15)                         | Non ( <i>n</i> = 15) | Cog ( <i>n</i> = 60)                                      | Non ( <i>n</i> = 60) |
| log(Freq)          | 1.7 (0.4)                                    | 1.7 (0.2)            | 1.7 (0.3)                                   | 1.7 (0.2)            | 1.7 (0.4)                                    | 1.7 (0.2)            | 1.8 (0.4)                                    | 1.8 (0.3)            | 1.7 (0.4)<br><i>t</i> (103.4) = -0.18, <i>p</i> = 0.86    | 1.7 (0.2)            |
| Length in phonemes | 5.0 (1.2)                                    | 5.1 (0.8)            | 5.1 (0.9)                                   | 4.8 (1.3)            | 4.6 (0.8)                                    | 5.0 (1.3)            | 5.3 (1.1)                                    | 5.2 (1.1)            | 5.0 (1.0)<br><i>t</i> (117.1) = -0.17, <i>p</i> = 0.87    | 5.0 (1.1)            |
| Eng phon nbors     | 4.3 (4.5)                                    | 4.5 (4.7)            | 5.1 (5.6)                                   | 5.1 (4.7)            | 4.5 (3.9)                                    | 4.5 (5.0)            | 4.9 (5.1)                                    | 3.9 (4.7)            | 4.7 (4.7)<br><i>t</i> (118.0) = 0.19, <i>p</i> = 0.85     | 4.5 (4.7)            |
| Spa phon nbors     | 0.00 (0.0)                                   | 0.00 (0.00)          | 0.20 (0.77)                                 | 0.00 (0.00)          | 0.00 (0.00)                                  | 0.07 (0.26)          | 0.13 (0.52)                                  | 0.00 (0.00)          | 0.08 (0.46)<br><i>t</i> (68.2) = 1.1, <i>p</i> = 0.29     | 0.02 (0.13)          |
| Mean phon freq     | 0.04 (0.02)                                  | 0.05 (0.01)          | 0.05 (0.02)                                 | 0.05 (0.01)          | 0.05 (0.02)                                  | 0.05 (0.01)          | 0.05 (0.02)                                  | 0.05 (0.02)          | 0.05 (0.02)<br><i>t</i> (113.4) = -0.23, <i>p</i> = 0.82  | 0.05 (0.01)          |
| Mean biphon freq   | 0.006 (0.004)                                | 0.006 (0.003)        | 0.005 (0.004)                               | 0.004 (0.004)        | 0.005 (0.004)                                | 0.004 (0.002)        | 0.006 (0.003)                                | 0.006 (0.005)        | 0.005 (0.004)<br><i>t</i> (118.0) = 0.51, <i>p</i> = 0.61 | 0.005 (0.004)        |
| Duration in ms     | 712 (143)<br><i>W</i> = 164, <i>p</i> = 0.03 | 615 (94)             | 738 (91)<br><i>W</i> = 160, <i>p</i> = 0.05 | 663 (174)            | 670 (119)<br><i>W</i> = 126, <i>p</i> = 0.59 | 655 (159)            | 706 (131)<br><i>W</i> = 132, <i>p</i> = 0.44 | 668 (147)            | 707 (122)<br><i>W</i> = 2,354, <i>p</i> = 0.004           | 651 (144)            |

Cognate vs. noncognate means did not differ for word frequency, length in phonemes, number of English or Spanish neighbors, or mean positional or biphone frequency (all  $|t|$ s < 1.7, all  $ps$  > 0.10), but Wilcoxon tests indicated stimulus durations were longer for cognates overall and in Set 1, and marginally so in Set 2.

versions of the experiments. The recordings were made in a sound-attenuated booth at a 44.1kHz sampling rate and 16-bit depth. The stimuli were later downsampled to 22,050Hz and converted to .mp3 format to minimize loading delays over the internet.

Speech-shaped noise was created by taking the long-term average spectrum of the files used to create the babble noise and then using the spectral shape as a filter for white noise; this was done with a Praat script derived from code used by Quené and Van Delft (2010). English and Spanish two-talker babble were created using freely available news podcasts. This allowed the choice of voices and accents while controlling for register and subject material across languages. Four podcasts were chosen, all with female presenters speaking a standard, not obviously regionally specific variety of the language. All non-speech noise, speech produced by a talker other than the main newscaster and pauses longer than 500ms were manually edited out. A random selection from each of the two English podcasts was then combined with each clip of target speech to make the English two-talker babble, and similarly for Spanish, using a Praat script (Boersma and Weenink, 2020).

The root mean square (RMS) amplitude for each clip of target speech was scaled to 70dB SPL, as was the noise, for a signal-to-noise ratio of 0dB. This ratio was chosen based on informal pilot testing and on findings in the literature suggesting that this level would be challenging but feasible for a range of participant profiles (Garcia Lecumberri et al., 2010).

## ANALYSIS AND RESULTS

### Overview of Analysis Procedures

All statistical analyses were carried out in R (R Core Team, 2020; version 4.0.3) using the following packages: lme4 (Bates et al., 2015; version 1.1.26), lmerTest (Kuznetsova et al., 2017; version 3.1.3), ggplot2 (Wickham, 2016; version 3.3.3), emmeans

(Lenth, 2021; version 1.6.3), sjPlot (Lüdtke, 2021; version 2.8.9), ggbiplot (Vu, 2011; version 0.55), and ggeffects (Lüdtke, 2018). Numerical predictors were centered and scaled, and transformed when appropriate, as determined by visual inspection of q-q plots. In general, the maximal random effects structure that would converge was used (Barr et al., 2013); more details regarding model selection are given below.

The first analysis asked whether the three participant groups showed differential effects of Cognate Status, Noise Condition, or their interaction. Unfortunately, due to model convergence issues (perhaps the result of quasi-separation; Kimball et al., 2019), it was not possible to fit a sufficiently complex mixed effects logistic regression to the accuracy data. We therefore present descriptive statistics for both real words and nonwords in order to qualitatively evaluate response strategies across groups and conditions. For example, particularly low nonword accuracy could indicate that some listeners were more likely to default to a “real word” response, potentially making it inappropriate to compare RTs across groups.

For the RT analysis, Participant Group was Helmert coded such that the first contrast compares the two bilingual groups to one another, and the second contrast compares the average of the bilingual groups to the monolinguals; any apparent cognate effects for the monolinguals are likely due to durational differences in the stimuli, so it is important to demonstrate that the bilinguals differ from the monolinguals. Stimulus Duration and its interactions with Noise Condition and Participant Group were also included as covariates. Both Cognate Status and Noise Condition were contrast coded; the coefficient for Cognate Status corresponds to the overall difference between cognates and noncognates, and the coefficients for Noise Condition provide comparisons of the following conditions: (1) Clear vs. all three noise conditions, indexing the overall recognition deficit in noise, (2) Speech-Shaped Noise vs. the average of English and Spanish Two-Talker Babble, indexing whether the groups responded

differently to energetic vs. informational masking, and (3) English Two-Talker Babble vs. Spanish Two-Talker Babble, indexing whether the effect of informational masking differed by masking language. The interaction terms involving Cognate Status, Participant Group, and Noise Condition thus ask whether the magnitude of any cognate effects was modulated by language background or masking noise.

The second set of analyses examined individual differences in RTs among the bilingual participants, treating language experience-related predictors as continuous rather than categorical, as has recently been advocated in the literature (Luk and Bialystok, 2013; Fricke et al., 2019). Principal component analysis was used to derive orthogonal measures of language experience, and model comparisons were used to determine the best-fitting model incorporating these measures.

## Group Analyses

### Descriptive Statistics for Word Recognition Accuracy

Trials with RTs faster than 300 ms were removed from the dataset (0.7% of the data). The cut-off point for long RTs was the one imposed by the experimental procedure (3,000 ms).

**Table 3** gives descriptive statistics for the accuracy data, first broken down more globally in terms of lexical status and the overall recognition deficit in noise (top portion), and second more granularly in terms of cognate status and noise condition (bottom portion). Accuracy for nonword stimuli can be taken as an index of word bias, with lower nonword accuracy indicating a greater tendency to default to a “real word” response.

There are several points to note. First, in the clear, word accuracy was at ceiling for all three participant groups at around 97%. Nonword accuracy was generally high but varied more than word accuracy: nonword accuracy for heritage bilinguals was 89%, versus 92% for L2 bilinguals and 94% for monolinguals. However, differences in both word and nonword accuracy were within 1 *SD* across groups.

With respect to masking effects, the three groups showed a comparable drop-off in word recognition accuracy of around 15–20%, with all groups again within a single *SD*. The noise deficit was numerically greatest for the heritage bilinguals, at 23% averaged across noise and cognate conditions. For all three groups, word biases increased considerably in noise, with nonword accuracy decreasing around 15–20% when averaged across all noise conditions. The numerically lowest nonword

accuracy was found for the L2 group in speech-shaped noise, at just 70%, though again this was within 1 *SD* of the other groups.

On average, cognates were recognized more accurately than noncognates in noise, by monolinguals as well as bilinguals, consistent with a slight potential benefit for longer stimuli in noise for all participant groups.

### Discussion of Accuracy Data

Word recognition accuracy suffered considerably in noise, and word biases increased substantially, with comparable effects across noise types and participant groups. In general, then, the three participant groups employed qualitatively similar response criteria and adjusted their response criteria in similar ways. Importantly, cognates tended to be recognized more accurately than noncognates, irrespective of participant language background. This suggests that the partial confound of stimulus duration and cognate status may be of concern and that the statistical model examining RTs should take this into account.

### Response Time Analysis

#### Data Preparation and Model Fitting Procedure

Response times (RTs) were measured from the onset of the target word and were log-transformed prior to analysis. All nonword trials, RTs faster than 300 ms, and incorrect responses (17% of the remaining data) were removed from the dataset. We then removed any responses that were more than two standard deviations from each participant's mean RT (4% of the remaining data). This data cleaning procedure left a total of 8,200 data points for analysis.

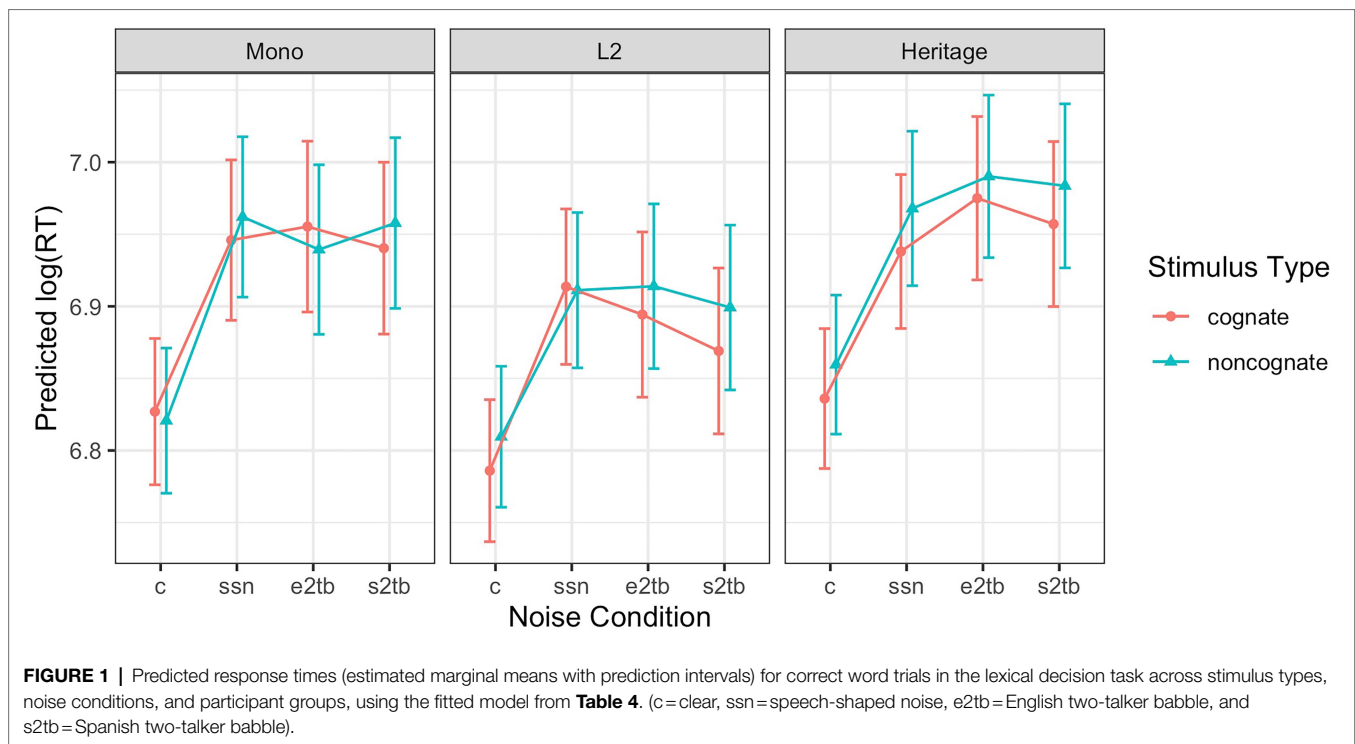
A model was fit that included all of the experimentally manipulated variables (Participant Group, Noise Condition, and Cognate Status) and their two- and three-way interactions as fixed effects, plus fixed effects of Stimulus Duration, and its two-way interactions with Participant Group and Noise Condition, plus the maximal random effects structure that would converge; this included by-participant and by-item random intercepts and slopes for the effect of Noise Condition.

### RT Results

**Figure 1** shows the predicted RT values for the fitted model, i.e., with effects of stimulus duration partialled out, and includes prediction intervals as implemented in the *ggeffects* R package. The model is given in **Table 4**, with coefficients numerically labeled for ease of reference.

**TABLE 3 |** Descriptive statistics for accuracy (proportion correct trials and *SD*s) by condition and participant group.

|                              | Monolingual English |             |             | L2 Spanish  |             |             | Heritage Spanish |             |             |
|------------------------------|---------------------|-------------|-------------|-------------|-------------|-------------|------------------|-------------|-------------|
|                              | Nonwords            | Words       |             | Nonwords    | Words       |             | Nonwords         | Words       |             |
| Overall accuracy (Clear)     | 0.94 (0.06)         | 0.97 (0.03) |             | 0.92 (0.10) | 0.96 (0.03) |             | 0.89 (0.12)      | 0.98 (0.03) |             |
| Overall accuracy (all noise) | 0.77 (0.09)         | 0.79 (0.09) |             | 0.73 (0.13) | 0.82 (0.08) |             | 0.75 (0.13)      | 0.75 (0.11) |             |
|                              | Nonwords            | Cog         | Noncog      | Nonwords    | Cog         | Noncog      | Nonwords         | Cog         | Noncog      |
| Clear                        | 0.94 (0.06)         | 0.97 (0.05) | 0.98 (0.03) | 0.92 (0.10) | 0.96 (0.05) | 0.97 (0.04) | 0.89 (0.12)      | 0.97 (0.04) | 0.98 (0.04) |
| SSN                          | 0.78 (0.12)         | 0.82 (0.11) | 0.73 (0.12) | 0.70 (0.15) | 0.84 (0.11) | 0.78 (0.14) | 0.73 (0.14)      | 0.79 (0.13) | 0.67 (0.14) |
| E2TB                         | 0.77 (0.16)         | 0.80 (0.14) | 0.79 (0.13) | 0.76 (0.12) | 0.83 (0.12) | 0.77 (0.12) | 0.75 (0.20)      | 0.79 (0.16) | 0.77 (0.17) |
| S2TB                         | 0.76 (0.12)         | 0.82 (0.19) | 0.77 (0.18) | 0.75 (0.15) | 0.87 (0.10) | 0.82 (0.13) | 0.76 (0.16)      | 0.77 (0.18) | 0.74 (0.21) |



In terms of main effects, Stimulus Duration was highly predictive of RT (1); longer stimuli elicited longer RTs. The Clear vs. All Noise comparison was significant (2) such that RTs in noise were slower than in the clear, and the L2 vs. Heritage listeners comparison was significant (3) such that RTs were faster for L2 as compared to Heritage listeners.

There were two two-way interactions involving Stimulus Duration, which are plotted as marginal effects in **Figure 2**. Stimulus Duration interacted with Noise Condition (4); the lengthening effect of longer duration on RTs was attenuated in noise as compared to in the clear. This was driven by disproportionately slower RTs to the shortest stimuli in noise (left panel). Stimulus Duration also interacted with Group (5); the lengthening effect of longer duration was attenuated for the Heritage listeners as compared to the L2 group. This was likewise driven by slower RTs to the shortest stimuli (right panel).

There was no main effect of Cognate Status, but the coefficient for Cognate Status differed for Bilinguals as compared to Monolinguals (6). Comparison of the estimated marginal means showed that the magnitude of the cognate effect was greater for Bilinguals as compared to Monolinguals (estimate=0.018, SE=0.008,  $p=0.02$ ). For Monolinguals for both cognates and noncognates, and for Bilinguals for noncognates only, the estimated mean log RT was 6.92; for Bilinguals for cognates, the mean was 6.90, corresponding to a cognate facilitation effect of about 20ms, all else being equal. Cognate Status did not enter into any additional interactions, indicating that the magnitude of the cognate effect was not modulated by the type of noise.

## Interim Discussion for RT Analyses

All listener groups responded more slowly in noise as compared to in the clear, and the overall degree of slowing was consistent across groups. While the former is expected, the latter is somewhat surprising in light of previous findings concerning heritage listeners' word recognition in noise (Mayo et al., 1997; Rogers et al., 2006; Morini and Newman, 2020). We consider this finding in more detail in the General Discussion.

The effects of cognate status on RT were surprisingly straightforward, though not as predicted. After statistically controlling for differences in stimulus duration, cognate facilitation was small in magnitude but significant for bilinguals as compared to monolinguals. Crucially, there was no evidence that the magnitude of facilitation was affected by the presence or type of competing noise. Contrary to what has been assumed in the literature, then, the results do not support the idea that cross-language activation is greater in noise relative to in the clear, at least during word recognition in the dominant language, a point we return to in the General Discussion. We next turn to the question of whether language experience modulated either the cognate effect or the noise masking effects.

## Individual Differences Analysis

### Overview of Individual Differences Analysis Procedure

The individual differences analysis asks whether differential experience in English vs. Spanish affects English word recognition, so we restrict our attention to the two bilingual groups. We focus on response times; the accuracy data suggested that the three groups employed similar response strategies, and the group

**TABLE 4** | Fixed and random effects for the model comparing RTs across participant groups.

| Label                 | Predictor                                           | Fixed effect estimates |      |             |              |
|-----------------------|-----------------------------------------------------|------------------------|------|-------------|--------------|
|                       |                                                     | Estimate               | SE   | CI (95%)    | p            |
|                       | (Intercept)                                         | 6.91                   | 0.01 | 6.88–6.94   | <0.001       |
| (1)                   | Stimulus duration                                   | 0.07                   | 0.01 | 0.06–0.09   | <0.001       |
| (2)                   | Noise Contrast 1 (Clear vs. All Noise)              | −0.12                  | 0.01 | −0.14–−0.10 | <0.001       |
|                       | Noise Contrast 2 (SSN vs. 2TB average)              | 0.00                   | 0.01 | −0.03–0.03  | 0.989        |
|                       | Noise Contrast 3 (E2TB vs. S2TB)                    | 0.01                   | 0.01 | −0.01–0.03  | 0.287        |
| (3)                   | Group Contrast 1 (L2 vs. Heritage)                  | 0.03                   | 0.02 | 0.00–0.06   | <b>0.034</b> |
|                       | Group Contrast 2 (Monoling vs. Biling)              | 0.00                   | 0.01 | −0.01–0.02  | 0.654        |
|                       | Cognate status                                      | 0.01                   | 0.01 | −0.01–0.04  | 0.302        |
| (4)                   | StimDur * NoiseC1                                   | 0.02                   | 0.01 | 0.01–0.03   | <b>0.003</b> |
|                       | StimDur * NoiseC2                                   | −0.00                  | 0.01 | −0.02–0.01  | 0.770        |
|                       | StimDur * NoiseC3                                   | −0.01                  | 0.01 | −0.02–0.01  | 0.320        |
| (5)                   | StimDur * GroupC1                                   | −0.00                  | 0.00 | −0.01–−0.00 | <b>0.021</b> |
|                       | StimDur * GroupC2                                   | 0.00                   | 0.00 | −0.00–0.00  | 0.160        |
|                       | NoiseC1 * CogStatus                                 | −0.00                  | 0.01 | −0.03–0.02  | 0.904        |
|                       | NoiseC2 * CogStatus                                 | −0.00                  | 0.01 | −0.03–0.03  | 0.946        |
|                       | NoiseC3 * CogStatus                                 | −0.02                  | 0.01 | −0.05–0.01  | 0.208        |
|                       | GroupC1 * CogStatus                                 | 0.00                   | 0.00 | −0.01–0.01  | 0.481        |
| (6)                   | GroupC2 * CogStatus                                 | −0.01                  | 0.00 | −0.01–−0.00 | <b>0.019</b> |
|                       | NoiseC1 * GroupC1                                   | −0.01                  | 0.01 | −0.03–0.01  | 0.429        |
|                       | NoiseC2 * GroupC1                                   | −0.02                  | 0.02 | −0.05–0.01  | 0.180        |
|                       | NoiseC3 * GroupC1                                   | −0.00                  | 0.01 | −0.02–0.02  | 0.694        |
|                       | NoiseC1 * GroupC2                                   | −0.00                  | 0.01 | −0.02–0.01  | 0.479        |
|                       | NoiseC2 * GroupC2                                   | 0.00                   | 0.01 | −0.02–0.02  | 0.763        |
|                       | NoiseC3 * GroupC2                                   | −0.01                  | 0.01 | −0.02–0.01  | 0.311        |
|                       | NoiseC1 * GroupC1 * CogStatus                       | −0.00                  | 0.01 | −0.02–0.01  | 0.655        |
|                       | NoiseC2 * GroupC1 * CogStatus                       | 0.02                   | 0.01 | −0.00–0.04  | 0.086        |
|                       | NoiseC3 * GroupC1 * CogStatus                       | −0.00                  | 0.01 | −0.02–0.02  | 0.972        |
|                       | NoiseC1 * GroupC2 * CogStatus                       | −0.01                  | 0.01 | −0.02–0.01  | 0.321        |
|                       | NoiseC2 * GroupC2 * CogStatus                       | 0.01                   | 0.01 | −0.00–0.02  | 0.193        |
|                       | NoiseC3 * GroupC2 * CogStatus                       | −0.01                  | 0.01 | −0.02–0.01  | 0.300        |
| <b>Random effects</b> |                                                     |                        |      |             |              |
|                       | $\sigma^2$                                          |                        |      | 0.02        |              |
|                       | $\tau_{00}$ Stimulus                                |                        |      | 0.01        |              |
|                       | $\tau_{00}$ Participant                             |                        |      | 0.01        |              |
|                       | $\tau_{11}$ Stimulus.NoiseC1                        |                        |      | 0.00        |              |
|                       | $\tau_{11}$ Stimulus.NoiseC2                        |                        |      | 0.00        |              |
|                       | $\tau_{11}$ Stimulus.NoiseC3                        |                        |      | 0.00        |              |
|                       | $\tau_{11}$ Participant.NoiseC1                     |                        |      | 0.01        |              |
|                       | $\tau_{11}$ Participant.NoiseC2                     |                        |      | 0.01        |              |
|                       | $\tau_{11}$ Participant.NoiseC3                     |                        |      | 0.00        |              |
|                       | Observations                                        |                        |      | 8,200       |              |
|                       | Marginal R <sup>2</sup> /Conditional R <sup>2</sup> |                        |      | 0.162/0.584 |              |

Significant fixed effects are labeled for ease of reference within the text.

analysis provided evidence that the RT model was able to adequately statistically control for the durational differences between stimulus types.

Seven language experience-related predictors were considered: composite self-ratings for Spanish proficiency (i.e., the averaged self-ratings for Spanish comprehension, speaking, and reading ability), Spanish age of acquisition, number of years in a Spanish-speaking household, self-reported current daily exposure to English and Spanish, and LexTALE scores in English and Spanish. English self-ratings, AoA, and years of household exposure were at or near ceiling, so they were not considered. **Table 5** gives Kendall's rank correlation coefficients for these seven variables; q-q plots indicated that all were non-normally

distributed to some extent. In general, the questionnaire responses were moderately intercorrelated, while the LexTALE scores were less correlated both with each other and with the self-report measures. To mitigate multicollinearity, principal component analysis was used to derive orthogonal indices of language experience.

### Principal Component Analysis

Principal Component Analysis was applied to the language experience predictors listed above. A Scree plot showed a drop-off in explained variance between the fifth and sixth principal components (PCs), so analysis was restricted to PCs one through five; each of these also accounted for at least 5%

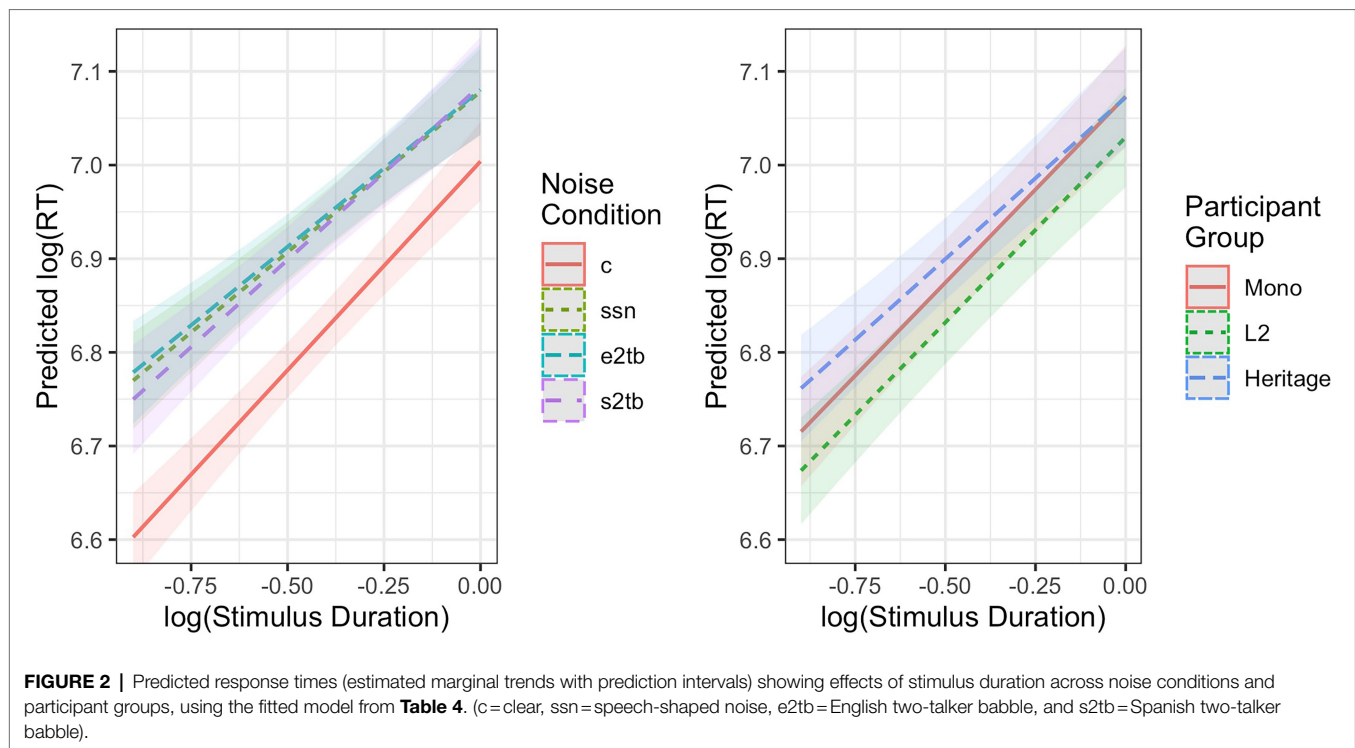


**TABLE 5 |** Correlation matrix (Kendall's  $\tau$ ) for individual difference measures.

|                        | Spa Composite Self-Rating            | Spa AoA                              | # Yrs in Spa Household               | % Daily Eng Exposure                 | % Daily Spa Exposure                | Eng LexTALE                        |
|------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|------------------------------------|
| Spa AoA                | -0.23, $p = 0.08$<br>$p$ adj. = 0.81 |                                      |                                      |                                      |                                     |                                    |
| # Yrs in Spa Household | 0.26, $p = 0.05$<br>$p$ adj. = 0.65  | -0.55, $p = 0.00$<br>$p$ adj. = 0.00 |                                      |                                      |                                     |                                    |
| % Daily Eng            | -0.31, $p = 0.02$<br>$p$ adj. = 0.33 | 0.28, $p = 0.03$<br>$p$ adj. = 0.53  | -0.28, $p = 0.03$<br>$p$ adj. = 0.52 |                                      |                                     |                                    |
| % Daily Spa            | 0.27, $p = 0.04$<br>$p$ adj. = 0.61  | -0.28, $p = 0.03$<br>$p$ adj. = 0.53 | 0.36, $p = 0.01$<br>$p$ adj. = 0.11  | -0.83, $p = 0.00$<br>$p$ adj. = 0.00 |                                     |                                    |
| Eng LexTALE            | -0.01, $p = 0.95$<br>$p$ adj. = 1.0  | 0.15, $p = 0.26$<br>$p$ adj. = 1.0   | -0.06, $p = 0.65$<br>$p$ adj. = 1.0  | 0.25, $p = 0.06$<br>$p$ adj. = 0.74  | -0.20, $p = 0.14$<br>$p$ adj. = 1.0 |                                    |
| Spa LexTALE            | 0.24, $p = 0.07$<br>$p$ adj. = 0.76  | -0.01, $p = 0.96$<br>$p$ adj. = 1.0  | 0.13, $p = 0.33$<br>$p$ adj. = 1.0   | -0.17, $p = 0.20$<br>$p$ adj. = 1.0  | 0.14, $p = 0.31$<br>$p$ adj. = 1.0  | 0.18, $p = 0.18$<br>$p$ adj. = 1.0 |

of variance, a value sometimes cited as a cut-off (Baayen, 2008). Subsequent analyses showed that PC5 did not predict RTs, so we do not consider it further. **Figure 3** plots each bilingual listener as a function of the first four PCs, which account for a total of 87% of the variance in language experience measures. **Table 6** gives the variable loadings for PCs 1–4; these correspond to the covariances between the seven original variables and the PCs.

PC1 (**Figure 3**, left panel) largely discriminates between the two bilingual groups. The heritage group is characterized by lower values of PC1, while the L2 group tends to have higher values; note that because the rotation of the PC axes is arbitrary, low values of PC1 correspond primarily

to greater daily Spanish exposure and more years in a Spanish-speaking household. We refer to this dimension as “PC1-DailySpaExposure.”

PC2 (left panel, y-axis) does not separate the participant groups and corresponds most closely to English LexTALE performance, followed by Spanish LexTALE performance. The heritage group shows greater variation along this axis than the L2 group. We refer to this dimension as “PC2-VerbalAbilityEng.”

PC3 (**Figure 3**, right panel) discriminates somewhat between participant groups and is most strongly associated with Spanish AoA, followed by number of years in a Spanish-speaking household. We refer to this dimension as “PC3-SpaAoA.”

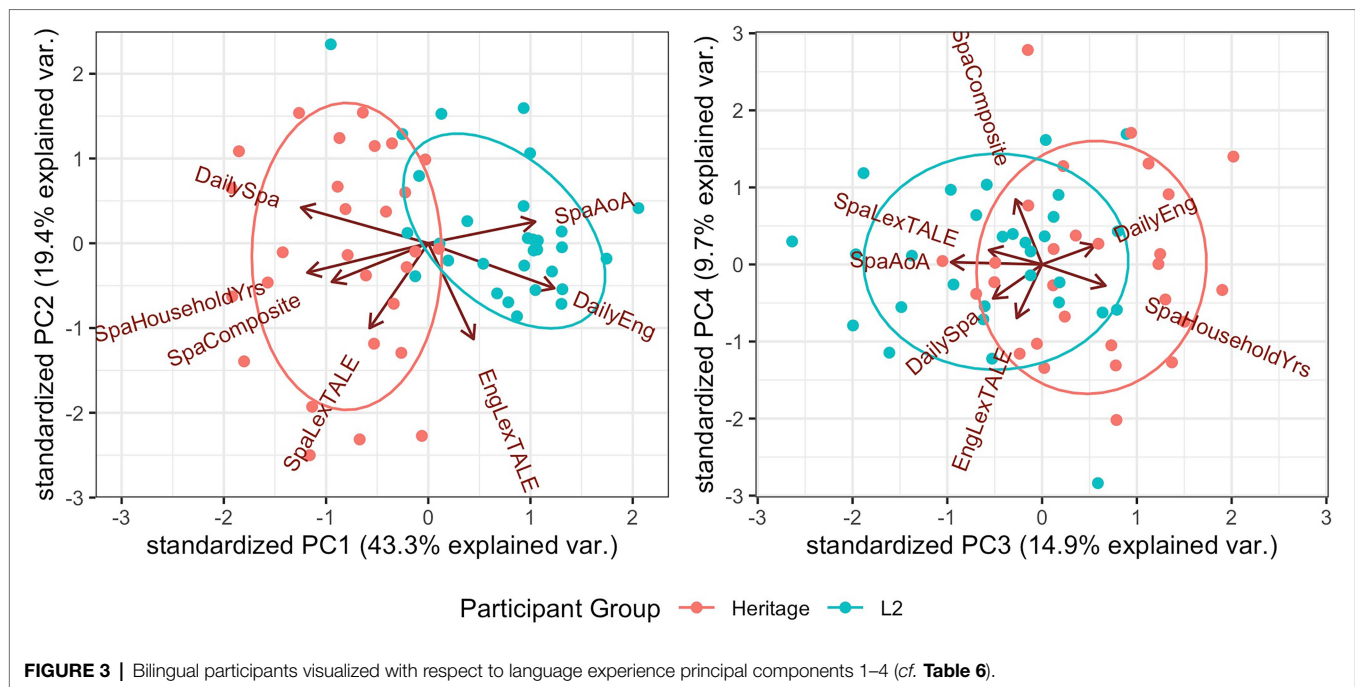


**TABLE 6 |** Variable loadings for PCs 1–4.

|                 | PC1 ("DailySpaExposure") | PC2 ("VerbalAbilityEng") | PC3 ("SpaAoA") | PC4 ("SpaSelfRating") |
|-----------------|--------------------------|--------------------------|----------------|-----------------------|
| SpaComposite    | −0.62                    | −0.30                    | −0.18          | <b>0.56</b>           |
| SpaAoA          | 0.69                     | 0.17                     | <b>−0.63</b>   | 0.02                  |
| SpaHouseholdYrs | <b>−0.78</b>             | −0.23                    | <b>0.44</b>    | −0.18                 |
| DailyEng        | <b>0.81</b>              | −0.35                    | 0.37           | 0.16                  |
| DailySpa        | <b>−0.82</b>             | 0.28                     | −0.34          | −0.29                 |
| EngLexTALE      | 0.29                     | <b>−0.75</b>             | −0.18          | <b>−0.46</b>          |
| SpaLexTALE      | −0.38                    | <b>−0.66</b>             | −0.37          | 0.12                  |

Variables mentioned in the text (i.e., those with highest covariances between variables and components) have been bolded.

Finally, PC4 (right panel, y-axis) corresponds primarily to Spanish self-ratings, with a smaller contribution from English LexTALE. Like PC2, PC4 does not separate the two bilingual groups, and the heritage group shows greater variation along this axis. We refer to this dimension as "PC4-SpaSelfRatings."

We note several additional considerations. First, we underscore that participants were overwhelmingly English dominant; all but four had higher self-ratings for English than Spanish, and all but two (different participants) had higher LexTALE scores in English than Spanish. The PCs derived from this analysis therefore index the strength of Spanish experience *relative to other English-dominant bilinguals who participated in the experiment*.

Second, Spanish LexTALE loaded weakly-to-moderately onto PCs 1–4, indicating that Spanish vocabulary scores did not represent a unique axis of variation in language experience for these bilinguals. However, based on previous work (e.g., Kilman et al., 2014; Scharenborg et al., 2018), we were nonetheless interested in the predictive power of objectively measured Spanish knowledge. We therefore included it in

the individual differences analyses alongside the PCs. Kendall's rank correlation tests with Holm-adjusted  $p$  values indicated that Spanish LexTALE scores were not significantly correlated with any of the PCs.

### Model Selection Procedure

A two-step model selection procedure was used. First, to determine whether any of the language experience measures predicted word recognition performance, we refit the full model from Table 4 to the bilingual RT data only, with Participant Group recoded as a two-way contrast between L2 and Heritage participants. In Table 7, this is referred to as the "base" model but note that model log likelihood varied depending on the maximal random effects structure that converged for each predictor. We fit four additional, nested models for each predictor: (1) the base model plus a main effect of the predictor, (2) model (1) plus the interaction of the predictor with Participant Group, (3) model (2) plus the interaction of the predictor with either Cognate Status or Noise Condition, and (4) model (3) plus the three-way interaction of the predictor with both

**TABLE 7 |** Nested comparisons of model log likelihood showing the predictive power of language experience predictors in the first step of the individual differences analysis.

| Significant cognate effect model comparisons                                             | logLik | $\chi^2$ | df | p       |
|------------------------------------------------------------------------------------------|--------|----------|----|---------|
| Base                                                                                     | 665.1  |          |    |         |
| Base + PC4                                                                               | 665.1  | 0.0      | 1  | 0.993   |
| Base + PC4 + PC4:Group                                                                   | 667.8  | 5.4      | 1  | 0.020   |
| Base + PC4 + PC4:Group + PC4:CogStatus                                                   | 669.3  | 2.9      | 1  | 0.087   |
| Base + PC4 + PC4:Group + PC4:CogStatus + PC4:CogStatus:Group                             | 671.2  | 3.9      | 1  | 0.048   |
| Significant noise effect model comparisons                                               | logLik | $\chi^2$ | df | p       |
| Base                                                                                     | 1595.0 |          |    |         |
| Base + PC1                                                                               | 1595.6 | 1.2      | 1  | 0.282   |
| Base + PC1 + PC1:Group                                                                   | 1596.8 | 2.3      | 1  | 0.125   |
| Base + PC1 + PC1:Group + PC1:NoiseCond                                                   | 1605.2 | 16.8     | 3  | < 0.001 |
| Base + PC1 + PC1:Group + PC1:NoiseCond + PC1:NoiseCond:Group                             | 1609.9 | 9.5      | 3  | 0.024   |
| Base                                                                                     | 1595.0 |          |    |         |
| Base + PC2                                                                               | 1595.1 | 0.1      | 1  | 0.790   |
| Base + PC2 + PC2:Group                                                                   | 1595.3 | 0.5      | 1  | 0.501   |
| Base + PC2 + PC2:Group + PC2:NoiseCond                                                   | 1598.5 | 6.4      | 3  | 0.093   |
| Base + PC2 + PC2:Group + PC2:NoiseCond + PC2:NoiseCond:Group                             | 1609.5 | 22.1     | 3  | < 0.001 |
| Base                                                                                     | 1885.0 |          |    |         |
| Base + PC4                                                                               | 1885.1 | 0.0      | 1  | 0.850   |
| Base + PC4 + PC4:Group                                                                   | 1885.2 | 0.2      | 1  | 0.668   |
| Base + PC4 + PC4:Group + PC4:NoiseCond                                                   | 1888.7 | 7.1      | 3  | 0.069   |
| Base + PC4 + PC4:Group + PC4:NoiseCond + PC4:NoiseCond:Group                             | 1891.6 | 5.8      | 3  | 0.124   |
| Base                                                                                     | 1595.0 |          |    |         |
| Base + SpaLexTALE                                                                        | 1595.1 | 0.1      | 1  | 0.782   |
| Base + SpaLexTALE + SpaLexTALE:Group                                                     | 1595.2 | 0.3      | 1  | 0.585   |
| Base + SpaLexTALE + SpaLexTALE:Group + SpaLexTALE:NoiseCond                              | 1603.2 | 15.9     | 3  | 0.001   |
| Base + SpaLexTALE + SpaLexTALE:Group + SpaLexTALE:NoiseCond + SpaLexTALE:NoiseCond:Group | 1604.9 | 3.5      | 3  | 0.327   |

Participant Group and Cognate Status/Noise Condition. This nested model evaluation procedure explicitly tests whether either the continuous language experience measures or the binary Group characterization add any predictive power over and above the other.

Random effects structures were adjusted following Barr et al. (2013) so that all nested models for a given predictor contained the same random effects. An  $\alpha$  level of 0.10 was used for model log likelihood comparisons; Matuschek et al. (2017) indicate that  $\alpha_{LRT}=0.10$  is somewhat more conservative than the criterion most commonly used for comparing model AICs, but not so conservative as to overly penalize model complexity. All significant model comparisons from this first step of the analysis procedure are given in **Table 7**; PCs 1, 2, and 4, and Spanish LexTALE improved model fit.

In the second step, we fit a saturated model consisting of the base model plus the significant predictors from the first step (i.e., PC1-DailySpaExposure, PC2-VerbalAbilityEng, PC4-SpaSelfRating, and Spanish LexTALE) and their two- and three-way interactions with Group and Noise Condition/Cognate Status. Backwards selection using nested model comparisons was then used to optimize model fit and statistical power (Matuschek et al., 2017). Leave-one-out comparisons indicated that all remaining interaction terms were significant at the  $\alpha_{LRT}=0.10$  level (all  $ps < 0.07$ ).

The final, fitted model is given in **Table 8**, with coefficients numerically labeled for ease of reference. The maximally converging random effects structure contained random by-participant and by-stimulus intercepts. In the text, we report pairwise comparisons

for the estimated marginal means at the average  $\pm 1SD$  of the values of each predictor, with familywise adjustments for  $p$  values as implemented in the emmeans package.

### Comparison of Individual Differences Model With Group Model

The model examining individual differences (IDs) among bilingual participants' RTs (**Table 8**) is qualitatively similar to the model comparing across all three participant groups (**Table 4**). Setting aside for a moment effects of the language experience predictors, the primary difference is that the IDs analysis returned an interaction between Noise Condition and Group such that the overall effect of noise (6) was greater for the Heritage as compared to the L2 group, and the effects of energetic vs. informational masking were reversed (7), such that the Heritage group suffered more in two-talker babble, while the L2 group suffered more in speech-shaped noise. This interaction is likely significant in the IDs model due to its less complex random effects structure; the group model contains random by-participant and by-stimulus slopes for the effect of Noise Condition, while the IDs model does not.

Turning to the language experience predictors, the coefficients for NoiseC2\*PC1 (8), NoiseC1\*PC4 (9), NoiseC2\*PC4 (10), NoiseC3\*SpaLexTALE (11), NoiseC1\*BilingGroup\*PC2, (13), and all three noise comparisons for BilingGroup\*PC4 (14–16) differed reliably from zero. The coefficients for CogStatus\*PC4 (12) and BilingGroup\*CogStatus\*PC4 (17) differed marginally from zero. **Figures 4, 5** visualize the prediction intervals involving these model coefficients.

**TABLE 8 |** Fixed and random effects for the model comparing RTs for bilingual participants only in the individual differences analysis.

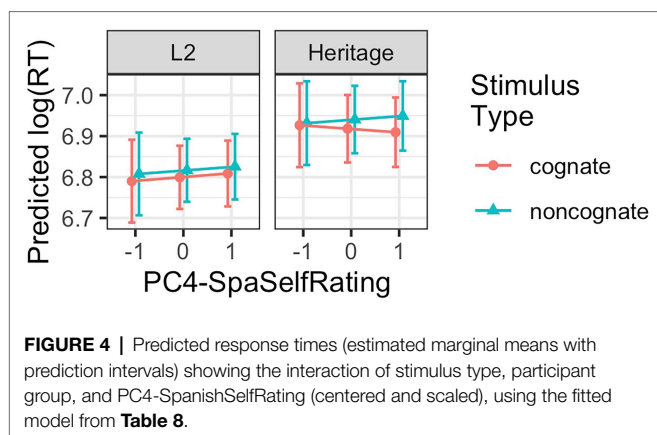
| Label                 | Coefficient                                         | Fixed effect estimates |             |             |                  |
|-----------------------|-----------------------------------------------------|------------------------|-------------|-------------|------------------|
|                       |                                                     | Estimate               | SE          | CI (95%)    | p                |
|                       | (Intercept)                                         | 6.87                   | 0.03        | 6.82–6.93   | <b>&lt;0.001</b> |
| (1)                   | Stimulus duration                                   | 0.07                   | 0.01        | 0.06–0.09   | <b>&lt;0.001</b> |
| (2)                   | Noise Contrast 1 (Clear vs. All Noise)              | –0.12                  | 0.01        | –0.13––0.11 | <b>&lt;0.001</b> |
|                       | Noise Contrast 2 (SSN vs. 2TB average)              | 0.00                   | 0.01        | –0.01–0.02  | 0.479            |
|                       | Noise Contrast 3 (E2TB vs. S2TB)                    | 0.01                   | 0.01        | –0.00–0.02  | 0.195            |
| (3)                   | BilingGroup                                         | 0.12                   | 0.06        | 0.01–0.24   | <b>0.033</b>     |
|                       | CogStatus                                           | 0.02                   | 0.01        | –0.01–0.05  | 0.180            |
|                       | PC1-DailySpaExposure                                | 0.05                   | 0.03        | –0.02–0.11  | 0.171            |
|                       | PC2-VerbalAbilityEng                                | 0.02                   | 0.02        | –0.03–0.07  | 0.419            |
|                       | PC4-SpaSelfRating                                   | 0.01                   | 0.02        | –0.03–0.04  | 0.735            |
|                       | SpaLexTALE                                          | 0.03                   | 0.03        | –0.03–0.08  | 0.335            |
| (4)                   | StimDur * NoiseC1                                   | 0.02                   | 0.01        | 0.01–0.03   | <b>&lt;0.001</b> |
|                       | StimDur * NoiseC2                                   | –0.01                  | 0.01        | –0.02–0.01  | 0.311            |
|                       | StimDur * NoiseC3                                   | –0.01                  | 0.01        | –0.02–0.00  | 0.188            |
| (5)                   | StimDur * BilingGroup                               | –0.01                  | 0.00        | –0.02––0.00 | <b>0.008</b>     |
|                       | NoiseC1 * CogStatus                                 | 0.00                   | 0.01        | –0.02–0.02  | 0.931            |
|                       | NoiseC2 * CogStatus                                 | –0.01                  | 0.01        | –0.03–0.02  | 0.476            |
|                       | NoiseC3 * CogStatus                                 | –0.02                  | 0.01        | –0.04–0.01  | 0.225            |
|                       | BilingGroup * CogStatus                             | 0.01                   | 0.01        | –0.01–0.03  | 0.514            |
| (6)                   | NoiseC1 * BilingGroup                               | –0.05                  | 0.02        | –0.09––0.01 | <b>0.007</b>     |
| (7)                   | NoiseC2 * BilingGroup                               | –0.08                  | 0.02        | –0.13––0.04 | <b>&lt;0.001</b> |
|                       | NoiseC3 * BilingGroup                               | –0.02                  | 0.03        | –0.07–0.04  | 0.553            |
|                       | BilingGroup * PC1                                   | –0.08                  | 0.05        | –0.18–0.03  | 0.139            |
|                       | NoiseC1 * PC1                                       | –0.02                  | 0.01        | –0.04–0.00  | 0.101            |
| (8)                   | NoiseC2 * PC1                                       | –0.03                  | 0.01        | –0.06––0.01 | <b>0.012</b>     |
|                       | NoiseC3 * PC1                                       | –0.01                  | 0.01        | –0.04–0.02  | 0.581            |
|                       | BilingGroup * PC2                                   | 0.01                   | 0.04        | –0.06–0.09  | 0.792            |
|                       | NoiseC1 * PC2                                       | 0.00                   | 0.01        | –0.01–0.02  | 0.696            |
|                       | NoiseC2 * PC2                                       | –0.02                  | 0.01        | –0.04–0.00  | 0.066            |
|                       | NoiseC3 * PC2                                       | –0.01                  | 0.01        | –0.03–0.01  | 0.373            |
|                       | BilingGroup * PC4                                   | –0.01                  | 0.03        | –0.07–0.06  | 0.834            |
| (9)                   | NoiseC1 * PC4                                       | –0.02                  | 0.01        | –0.03––0.01 | <b>&lt;0.001</b> |
| (10)                  | NoiseC2 * PC4                                       | –0.04                  | 0.01        | –0.05––0.03 | <b>&lt;0.001</b> |
|                       | NoiseC3 * PC4                                       | 0.01                   | 0.01        | –0.00–0.03  | 0.094            |
|                       | NoiseC1 * SpaLexTALE                                | –0.01                  | 0.01        | –0.02–0.01  | 0.545            |
|                       | NoiseC2 * SpaLexTALE                                | –0.02                  | 0.01        | –0.04–0.01  | 0.171            |
| (11)                  | NoiseC3 * SpaLexTALE                                | –0.04                  | 0.01        | –0.06––0.01 | <b>0.002</b>     |
| (12)                  | CogStatus * PC4                                     | 0.01                   | 0.00        | –0.00–0.02  | 0.070            |
|                       | NoiseC1 * BilingGroup * CogStatus                   | –0.02                  | 0.02        | –0.06–0.02  | 0.382            |
|                       | NoiseC2 * BilingGroup * CogStatus                   | 0.04                   | 0.02        | –0.01–0.08  | 0.131            |
|                       | NoiseC3 * BilingGroup * CogStatus                   | –0.00                  | 0.03        | –0.06–0.05  | 0.861            |
| (13)                  | (NoiseC1 * BilingGroup) * PC2                       | –0.05                  | 0.01        | –0.07––0.02 | <b>&lt;0.001</b> |
|                       | (NoiseC2 * BilingGroup) * PC2                       | 0.00                   | 0.01        | –0.03–0.03  | 0.936            |
|                       | (NoiseC3 * BilingGroup) * PC2                       | 0.01                   | 0.02        | –0.03–0.04  | 0.726            |
| (14)                  | (NoiseC1 * BilingGroup) * PC4                       | –0.03                  | 0.01        | –0.06––0.01 | <b>0.003</b>     |
| (15)                  | (NoiseC2 * BilingGroup) * PC4                       | –0.03                  | 0.01        | –0.05––0.00 | <b>0.047</b>     |
| (16)                  | (NoiseC3 * BilingGroup) * PC4                       | 0.04                   | 0.02        | 0.01–0.07   | <b>0.004</b>     |
| (17)                  | (BilingGroup * CogStatus) * PC4                     | 0.02                   | 0.01        | –0.00–0.04  | 0.055            |
| <b>Random effects</b> |                                                     |                        |             |             |                  |
|                       | $\sigma^2$                                          |                        | 0.03        |             |                  |
|                       | $\tau_{00}$ Stimulus                                |                        | 0.01        |             |                  |
|                       | $\tau_{00}$ Participant                             |                        | 0.01        |             |                  |
|                       | N Participant                                       |                        | 58          |             |                  |
|                       | N Stimulus                                          |                        | 116         |             |                  |
|                       | Observations                                        |                        | 5,369       |             |                  |
|                       | Marginal R <sup>2</sup> /Conditional R <sup>2</sup> |                        | 0.188/0.505 |             |                  |

Significant fixed effects are labeled for ease of reference within the text.

### Individual Differences in the Cognate Effect

Similar to the group analysis, there was no main effect of Cognate Status and no interaction between Cognate Status

and Noise Condition. **Figure 4** depicts the interaction of Cognate Status, PC4-SpaSelfRating, and Group. For L2 Spanish listeners, the difference in RTs between cognates and noncognates remained



constant across values of PC4-SpaSelfRating and was not statistically significant (estimate =  $-0.02$ ,  $SE = 0.02$ ,  $p = 0.28$ ). For Heritage listeners, the difference between cognates and noncognates was nonexistent at lower values of PC4-SpaSelfRating (for the lowest values, estimate =  $-0.01$ ,  $SE = 0.02$ ,  $p = 0.91$ ; for average values, estimate =  $-0.02$ ,  $SE = 0.02$ ,  $p = 0.22$ ), and significant at the highest values (estimate =  $-0.04$ ,  $SE = 0.02$ ,  $p = 0.03$ ).

### Individual Differences in Masking Effects

The lower left panel of Figure 5 depicts the interaction of Noise Condition and PC1-DailySpaExposure (8), averaging across listener groups and the two-talker babble conditions in line with the model. On average, RTs in the clear were slower for high vs. low values of PC1-DailySpaExposure, but this difference was not significant (estimate =  $0.06$ ,  $SE = 0.07$ ,  $p = 0.35$ ). The recognition deficit for the two-talker babble conditions (i.e., the difference in RTs between the 2TB and clear conditions) was smaller at low values of PC1-DailySpaExposure as compared to high values (estimated difference =  $0.06$ ,  $SE = 0.02$ ,  $p = 0.03$ ), while the recognition deficit for speech-shaped noise remained constant across values of PC1-DailySpaExposure (estimated difference =  $-0.01$ ,  $SE = 0.03$ ,  $p = 0.95$ ).

The lower right panel of Figure 5 depicts the two-way interaction of Noise Condition with SpaLexTALE (11). Across listener groups, the recognition deficit in S2TB increased along with SpaLexTALE scores (estimate =  $0.06$ ,  $SE = 0.02$ ,  $p = 0.03$ ), while recognition deficits remained constant for both the SSN (estimate =  $-0.01$ ,  $SE = 0.02$ ,  $p = 0.97$ ) and E2TB conditions (estimate =  $-0.02$ ,  $SE = 0.02$ ,  $p = 0.83$ ).

The upper left panel of Figure 5 depicts the three-way interaction of Group and Noise Condition with PC2-VerbalAbilityEng (14), with the three noise conditions averaged together in line with the statistical model; PC2-VerbalAbilityEng did not interact with masker type. For L2 listeners, RTs in the clear tended to increase along with PC2-VerbalAbilityEng, while for Heritage listeners, RTs in noise tended to increase along with PC2-VerbalAbilityEng, though neither of these trends were significant (for L2, estimate =  $0.07$ ,  $SE = 0.07$ ,  $p = 0.51$ ; for Heritage, estimate =  $0.06$ ,  $SE = 0.06$ ,  $p = 0.50$ ). However, L2 listeners showed comparatively smaller noise deficits

at high (vs. low) values of PC2 (estimate =  $-0.05$ ,  $SE = 0.02$ ,  $p = 0.02$ ), while Heritage listeners trended toward the opposite pattern (estimate =  $0.04$ ,  $SE = 0.02$ ,  $p = 0.07$ ).

The upper right panel of Figure 5 depicts the three-way interaction of Group, Noise Condition, and PC4-SpaSelfRating (14–16). For L2 listeners, the recognition deficits in all three types of noise remained constant across values of PC4-SpaSelfRating (for SSN, estimate =  $-0.03$ ,  $SE = 0.02$ ,  $p = 0.45$ ; for E2TB, estimate =  $0.01$ ,  $SE = 0.02$ ,  $p = 0.81$ ; and for S2TB, estimate =  $0.04$ ,  $SE = 0.02$ ,  $p = 0.18$ ). For Heritage listeners, recognition deficits increased along with PC4-SpaSelfRating in E2TB (estimate =  $0.14$ ,  $SE = 0.02$ ,  $p < 0.001$ ) and S2TB (estimate =  $0.08$ ,  $SE = 0.02$ ,  $p < 0.001$ ), while the deficit in SSN remained constant (estimate =  $0.01$ ,  $SE = 0.02$ ,  $p = 0.97$ ). Moreover, the difference in the recognition deficit for high vs. low levels of PC4-SpaSelfRating was greater for E2TB as compared to S2TB (estimate =  $0.07$ ,  $SE = 0.02$ ,  $p = 0.001$ ).

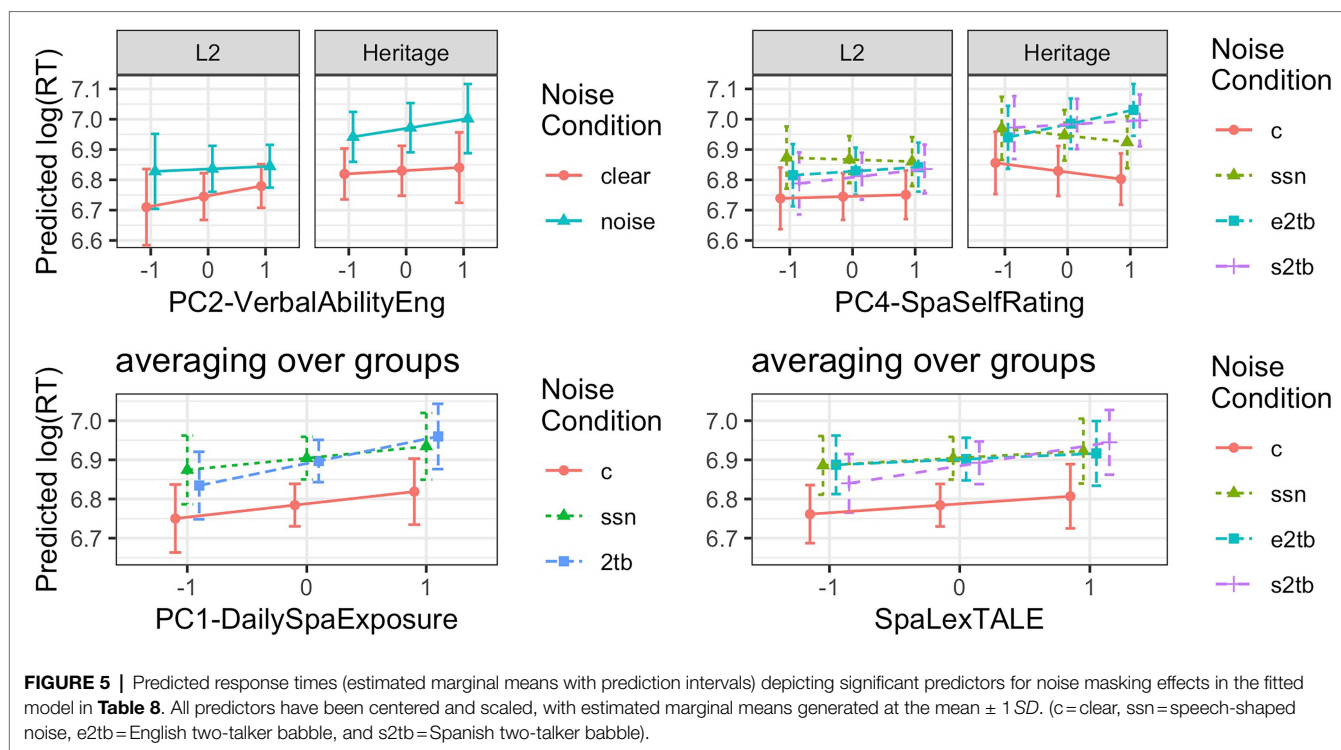
### Interim Discussion for Individual Differences Analyses

We explored the effects of bilingual language experience on English word recognition in noise using principal component analysis. Three PCs (PC1-DailySpaExposure, PC2-VerbalAbilityEng, and PC4-SpaSelfRating) and the Spanish LexTALE scores improved the model's ability to predict the masking effects, and PC4-SpaSelfRating also improved the model's ability to predict the cognate effect.

For both bilingual listener groups, PC1-DailySpaExposure predicted the degree of informational masking, with lower PC1-DailySpaExposure values (i.e., greater daily Spanish exposure) associated with less masking from two-talker babble as compared to higher values of PC1-DailySpaExposure. Given that all participants were English-dominant, this result suggests that regular exposure to a non-dominant language may confer benefits in terms of coping with informational masking; we return to this idea in the General Discussion. The modulation in masking effects associated with Spanish LexTALE was also consistent across listener groups: bilinguals with higher Spanish LexTALE scores experienced more disruption from Spanish two-talker babble as compared to bilinguals with lower scores.

The effects of PC2-VerbalAbilityEng differed by listener group. For L2 listeners, lower values of PC2-VerbalAbilityEng (associated with a larger English vocabulary size) were associated with faster English word recognition in the clear, while for heritage listeners, lower values of PC2-VerbalAbilityEng (i.e., larger English vocabulary size) were associated with marginally reduced word recognition difficulties in noise. The fact that the effects of PC2-VerbalAbilityEng differed across groups indicates that English LexTALE scores index something different in L1 English-L2 Spanish listeners vs. heritage listeners, a point we return to in "Implications for Measures of Language Experience."

PC4 corresponded most closely to Spanish self-ratings, with higher PC4 values corresponding to higher self-ratings. For L2 Spanish listeners, recognition deficits in noise were not modulated by PC4-SpaSelfRating. For heritage listeners, the



deficits incurred in two-talker babble were greater for listeners with higher values of PC4-SpaSelfRating, and this effect was larger for English two-talker babble as compared to Spanish two-talker babble. This finding suggests that for heritage listeners, Spanish self-ratings may have actually been a better index of past English experience; we discuss this possibility in “Language Experience and Masking Effects.”

Finally, model comparisons returned a three-way interaction of PC4-SpaSelfRating, Participant Group, and Cognate Status. Pairwise comparisons of the estimated marginal means indicated that for L2 listeners, cognate effects did not differ reliably from zero, irrespective of PC4-SpaSelfRating. For heritage listeners, cognate effects differed reliably from zero only for those with the highest values of PC4-SpaSelfRating. Importantly, however, there was still no indication that cognate facilitation effects were modulated by the presence of background noise.

## GENERAL DISCUSSION

### No Evidence for Increased Cross-Language Activation in Noise

In “Mechanisms by Which Cross-Language Activation Could Increase in Noise,” we hypothesized that one possible mechanism for increased cross-language activation in noise could be phonetically driven changes in the competitor activation process. Because competing noise makes the target phonetic input less recoverable (e.g., Bradlow and Alexander, 2007) and less reliable (e.g., McQueen and Huettig, 2012), non-target language competitors could be activated more strongly during

the recognition process in noise than in the clear, leading to a larger cognate effect in noise.

Contrary to this proposal, cognate facilitation in this study did not increase in noise. This finding is striking in that it seems to run counter to assumptions made in the literature (e.g., Rogers et al., 2006). However, because this study examined performance in the more dominant language only, further studies should investigate whether cross-language activation increases in noise during word recognition in the less dominant language. Given the lack of support for the phonetically driven hypothesis, the present results suggest that if cross-language activation processes are altered by noise in the non-dominant language, factors involving executive function, language control, and the availability of cognitive resources will likely be responsible.

## Relating Language Experience to Auditory Word Recognition in Noise

### Language Experience and Cognate Facilitation

While background noise did not impact cross-language activation, language experience did play a role. The RT analysis across all three listener groups showed that relative to the monolingual control group, bilinguals experienced cognate facilitation during English auditory word recognition. The individual differences analyses indicated that cognate facilitation was weak-to-nonexistent for L2 Spanish listeners and strongest for heritage listeners with the highest Spanish self-ratings. The two analyses may have differed in their estimation of the cognate effect for several reasons. The group analysis included a more complex

random effect structure that was better able to account for stimulus duration and random by-participant differences, and the monolingual control comparison also helped account for sources of variation not attributable to bilingualism.

Interestingly, the magnitude of the cognate effect for heritage bilinguals was associated primarily with differences in Spanish self-ratings, and not with Spanish LexTALE scores. Since Spanish LexTALE scores predicted the degree of masking from Spanish two-talker babble, LexTALE seems to have served as a reasonable approximation of Spanish proficiency. One interpretation of the cognate finding is therefore that heritage bilinguals' Spanish self-ratings reflected some latent aspect(s) of English exposure, an interpretation supported by the fact that English LexTALE scores also loaded onto PC4-SpaSelfRating. If heritage listeners with the highest Spanish self-ratings tended to have the weakest English lexical-phonetic representations, then we might predict them to show (1) slower RTs for noncognates, which do not benefit from cross-language activation and/or (2) faster RTs for cognates, which may be more strongly influenced by cross-language activation for listeners with the weakest representations. While there were weak trends in these directions, the cognate analysis itself does not enable us to say more on this point. However, the analysis of masking effects is relevant here, and we return to this issue below.

### Language Experience and Masking Effects

The present study is one of relatively few to directly compare the effects of different noise types (see Scharenborg and van Os, 2019, for a recent review). Contrary to some findings (e.g., Cooke et al., 2010; Kilman et al., 2014), the group analysis here indicated that for most listeners, competing speech was no more disruptive than stationary noise, although the individual differences analyses painted a more nuanced picture. Importantly, only the group analysis incorporated random by-participant slopes for the effect of noise type. Such a model is likely to attribute differential effects of noise type to random variation in the listener population, when in fact they are partially explicable by systematic differences in language experience. Indeed, the individual differences analysis showed that differences in daily language exposure, Spanish proficiency, and English verbal ability all helped predict inter-individual differences in masking effects.

For both bilingual groups, greater daily Spanish exposure (corresponding to lower values of PC1-DailySpaExposure; see **Figure 3**) was associated with an improved ability to cope with informational masking. Interestingly, this finding identifies the quantity of non-dominant language exposure as the operative factor, and *not* proficiency in either the target or masking languages. This finding is compatible with the proposal that experience regulating the more dominant language, i.e., more time spent listening in the non-dominant language, hones the deployment of cognitive resources (e.g., Alladi et al., 2013). It also supports the argument that research examining relationships among bilingual language experience and cognitive function should move beyond static (i.e., proficiency-oriented) measures to focus more on dynamic measures of language experience (e.g., Beatty-Martínez et al., 2020). The question

of *which aspects* of daily language experience may promote an improved ability to cope with informational masking should therefore be a topic of future research.

That greater daily Spanish exposure was associated with a reduced noise deficit seems to run counter to von Hapsburg and Bahng (2009), who found that L2 immersion was associated with more impaired L1 word recognition in noise. Several differences between studies should be noted here. First, the current listeners were not in any sense fully immersed in the non-dominant language. The proportion of daily Spanish exposure averaged 0.19 and ranged from 0.01 to 0.57; contextual support for the dominant language was thus quite high. Second, the modulation of the noise deficit found here was restricted to the two-talker babble conditions, indicating a relationship specifically between non-dominant language exposure and the ability to cope with informational masking; von Hapsburg and Bahng examined only energetic masking. Future research should further explore the relationships among non-dominant language exposure, informational masking, and domain-general executive function.

In addition to the effect of daily Spanish exposure on informational masking, Spanish proficiency (i.e., Spanish LexTALE scores) predicted the degree of informational masking incurred by Spanish babble, an effect that was equivalent across the two bilingual groups. This finding is roughly in line with previous SPIN findings (Imai et al., 2005; Van Engen, 2010; Brouwer et al., 2012; Kilman et al., 2014), but it extends these in several ways; namely, by moving beyond group-level analyses to relate the objectively measured vocabulary knowledge of individual listeners to the magnitude of the interference effect, and also by identifying such an effect during word recognition in the dominant language.

With respect to effects of English language experience, heritage listeners' overall noise deficits were marginally related to PC2-VerbalAbilityEng, which was primarily composed of English LexTALE scores. Informational masking was also greater overall for heritage listeners as compared to the L2 Spanish group, and it increased along with values of PC4-SpaSelfRating. As alluded to previously, the fact that PC4-SpaSelfRating predicted informational masking independently of Spanish LexTALE scores suggests that PC4-SpaSelfRating reflected some aspect of English experience. Importantly, increasing values of PC4-SpaSelfRating were associated with more sharply increasing disruption from competing English speech as compared to competing Spanish speech (**Figure 5**, upper right); this suggests that the relative weakness of the target English representations may be most relevant here, and not the strength of competing Spanish representations. Under this interpretation, listeners with the weakest English representations were the most susceptible to informational masking, and informational masking was strongest when the target and competing speech were most similar (Van Engen, 2010). The same listeners were also most likely to show cognate facilitation effects. Taken together, the results suggest that (1) for heritage listeners, PC4-SpaSelfRating

likely indexed some aspect of early-acquired English lexical-phonetic knowledge and (2) this knowledge was distinct from daily language exposure (i.e., PC1-DailySpaExposure), Spanish age of acquisition (PC3-SpaAoA), or Spanish vocabulary knowledge (SpaLexTALE). This in turn suggests that cognate facilitation for heritage listeners was likewise driven by differences in early-acquired English lexical-phonetic knowledge that were not captured by these other metrics.

Finally, returning to the lack of differentiation among noise conditions in the group analysis, differences in task difficulty across studies may have also played a role. Calandruccio et al. (2010) argued that effects of informational masking were greatest when the speech comprehension system was most stressed, i.e., when task difficulty was highest. Previous studies employed tasks that may have been more cognitively taxing than the present experiment; Cooke et al. (2010) used a 24AFC consonant identification task, and Kilman et al. (2014) used a sentence repetition paradigm. The group analysis may therefore have found no differences across noise types because lexical decision in the native language does not generally incur enough cognitive load to reveal significant effects of informational masking. The findings from the individual differences analysis involving heritage listeners support this interpretation; the combination of relatively weaker English lexical-phonetic representations and relatively stronger competition from Spanish representations is likely to have increased task difficulty for some heritage listeners, making them particularly susceptible to informational masking.

## Implications for Measures of Language Experience

Principal component analysis suggested four main axes along which the English-dominant bilinguals in this study varied: daily exposure to Spanish vs. English (PC1-DailySpaExposure; 43% of the total variance in language experience measures); verbal ability primarily in the more dominant language, English (PC2-VerbalAbilityEng; 19% of variance); Spanish age of acquisition (PC3-SpaAoA; 15% of variance); and self-rated proficiency in the non-dominant language, Spanish (PC4-SpaSelfRating; 10% of variance). Three of these helped explain variance in English word recognition in noise; age of acquisition, notably, did not. That measures capturing more nuanced aspects of accumulated experience were predictive, and not AoA *per se*, is broadly consistent with perspectives emphasizing the role of plasticity and continued learning over the life span (Baum and Titone, 2014; Flege and Bohn, 2021).

The fact that Spanish LexTALE was not uniquely associated with any single PC, but rather loaded into multiple components, indicates that receptive vocabulary knowledge in the non-dominant language was not a unique axis of variation for the bilinguals in this study. However, Spanish LexTALE scores predicted the degree of masking incurred by competing Spanish speech. Given previous findings that LexTALE performance correlates with other aspects of proficiency in the non-dominant language (Lemhöfer and Broersma, 2012),

Spanish LexTALE likely predicted Spanish masking because it indexed listeners' facility in deriving meaning from competing Spanish speech.

However, the fact that PC2-VerbalAbilityEng (composed predominantly of English LexTALE scores) yielded different effects for L2 vs. heritage listeners indicates that LexTALE scores may only provide a useful measure of dominant-language linguistic knowledge for certain populations of bilinguals. If L2 listeners' English language knowledge was more likely to be at ceiling, then performance in the two lexical decision tasks (LexTALE and SPIN) may have reflected individual differences in processing speed, attentional focus, or other domain-general attributes. On the other hand, heritage listeners' performance in the two lexical decision tasks may have tended to reflect variation in English lexical knowledge. If correct, this adds nuance to Ferré and Brysbaert's (2017) suggestion that LexTALE can discriminate among bilinguals at the high end of the proficiency range even for the more dominant language; LexTALE may specifically provide an appropriate proficiency measure only for bilinguals whose dominance has shifted over time, or perhaps for bilinguals whose language input has been more equally shared across languages over the life span.

The effects of PC4-SpaSelfRating also differed across groups, aligning with recent demonstrations that self-ratings can reflect different aspects of experience for different populations (Tomoschuk et al., 2019). The results therefore support calls to incorporate objective language proficiency measures (Kilman et al., 2014; Warzybok et al., 2015; Scharenborg et al., 2018), with the caveats that (1) caution is warranted in using LexTALE as a measure of linguistic knowledge in the more dominant language and (2) self-reported measures still have an important role to play (see Gollan et al., 2011b and Gullifer et al., 2021, for more nuanced discussion).

While the present work has begun to identify how different aspects of linguistic experience impact bilingual SPIN, a more complete understanding will require identifying how experience impacts specific components of the recognition process (see Beatty-Martínez and Titone, 2021 and Green and Wei, 2014 for similar arguments regarding executive function). Work by Krizman et al. (2017) has begun separating out the relevant components of auditory-linguistic processing. Future work should ideally merge these streams of inquiry.

## Implications for Heritage Speaker Populations

The present results add nuance to previous findings regarding speech perception in heritage speaker populations. On average, heritage listeners were no slower than monolinguals to recognize words in English, though they were more affected by the presence of noise and by informational masking in particular as compared to L1 English-L2 Spanish listeners. The results therefore align with previous literature in suggesting that heritage listeners' lexical-phonetic knowledge in the more dominant language may in some cases be less robust as compared to listeners who acquired

a single language in childhood (Mayo et al., 1997; Rogers et al., 2006; Morini and Newman, 2020). However, they also demonstrate that in some cases, such effects may be small-to-negligible. Future studies should therefore include large sample sizes to guard against inappropriately concluding that heritage listeners are uniformly disadvantaged in word recognition. They should also combine careful and detailed measures of language experience in order to further disentangle the factors influencing language processing behavior in this population.

## CONCLUSION

In addition to being one of relatively few studies to report cognate activation during auditory word recognition (Woutersen et al., 1995; Blumenfeld and Marian, 2007), this study showed that cross-language activation processes were not affected by background noise during word recognition in the dominant language. A detailed exploration of individual differences indicated that the ability to cope with informational masking was particularly subject to modulation by language experience. Taken together, the findings confirm the highly interactive nature of bilingual language processing and suggest that auditory word recognition processes in the native language remain susceptible to influence from linguistic experience throughout the lifespan.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## REFERENCES

- Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., et al. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology* 81, 1938–1944. doi: 10.1212/01.wnl.0000436620.33155.a4
- Antoniou, M. (2019). The advantages of bilingualism debate. *Annu. Rev. Linguist.* 5, 395–415. doi: 10.1146/annurev-linguistics-011718-011820
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behav. Res. Methods* 52, 388–407. doi: 10.3758/s13428-019-01237-x
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. United Kingdom: Cambridge University Press.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Baum, S., and Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Appl. Psycholinguist.* 35, 857–894. doi: 10.1017/S0142716414000174
- Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., and Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 1022–1047. doi: 10.1037/xlm0000770

## ETHICS STATEMENT

This study involving human participants was reviewed and approved by the Human Research Protection Office, University of Pittsburgh. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

All aspects of the study were carried out by MF (study design, stimulus preparation, data collection, analysis, and manuscript preparation) with help from research assistants as noted in the Acknowledgments.

## FUNDING

This research was supported by a Language Learning Early Career Grant to MF.

## ACKNOWLEDGMENTS

The author gratefully acknowledges the editor and reviewers for helpful and constructive feedback that significantly improved the paper. Many thanks go to Sofia Palacios; to Cecilia Jimenez and Ivy Yen for help in compiling stimuli; and thanks to Cecilia, Ivy, Alicia Sigmon, Melinda Arnold, and Miroo Lee for helpful discussion. Thanks also to Rachel Theodore for advice on implementing the headphone screening task and for her very helpful tutorial on using Gorilla for online speech perception experiments.

- Beatty-Martínez, A. L., and Titone, D. A. (2021). The quest for signals in noise: leveraging experiential variation to identify bilingual phenotypes. *Languages* 6:168. doi: 10.3390/languages6040168
- Bice, K., and Kroll, J. F. (2021). Grammatical processing in two languages: how individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals. *J. Neurolinguistics* 58:100963. doi: 10.1016/j.jneuroling.2020.100963
- Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Front. Psychol.* 9:81. doi: 10.3389/fpsyg.2018.00081
- Black, J. W., and Hast, M. H. (1962). Speech reception with altering signal. *J. Speech Hear. Res.* 5, 70–75. doi: 10.1044/jshr.0501.70
- Blumenfeld, H. K., and Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: examining proficiency and lexical status using eye-tracking. *Lang. Cogn. Process.* 22, 633–660. doi: 10.1080/01690960601000746
- Blumenfeld, H. K., and Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *J. Cogn. Psychol.* 25, 547–567. doi: 10.1080/20445911.2013.812093
- Boersma, P., and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.09. Available at: <http://www.praat.org> (Accessed February 14, 2020).
- Bolger, P. A., and Zapata, G. C. (2011). Psycholinguistic approaches to language processing in heritage speakers. *Heritage Lang. J.* 8, 1–29. doi: 10.46538/hlj.8.1.2
- Bradlow, A. R., and Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J. Acoust. Soc. Am.* 121, 2339–2349. doi: 10.1121/1.2642103

- Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., et al. (2001). Context processing in older adults: evidence for a theory relating cognitive control to neurobiology in healthy aging. *J. Exp. Psychol. Gen.* 130, 746–763. doi: 10.1037/0096-3445.130.4.746
- Broersma, M., and Cutler, A. (2011). Competition dynamics of second-language listening. *Q. J. Exp. Psychol.* 64, 74–95. doi: 10.1080/17470218.2010.499174
- Brouwer, S., and Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. *J. Psycholinguist. Res.* 45, 1151–1160. doi: 10.1007/s10936-015-9396-9
- Brouwer, S., Van Engen, K. J., Calandruccio, L., and Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *J. Acoust. Soc. Am.* 131, 1449–1464. doi: 10.1121/1.3675943
- Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Calandruccio, L., Dhar, S., and Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J. Acoust. Soc. Am.* 128, 860–869. doi: 10.1121/1.3458857
- Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., and Wagner, K. (2010). Carpet or Cárcel: the effect of age of acquisition and language mode on bilingual lexical access. *Lang. Cogn. Process.* 25, 669–705. doi: 10.1080/01690960903474912
- Caramazza, A., and Brunes, I. (1979). Lexical access in bilinguals. *Bull. Psychon. Soc.* 13, 212–214. doi: 10.3758/BF03335062
- Chambers, C. G., and Cooke, H. (2009). Lexical competition during second-language listening: sentence context, but not proficiency, constrains interference from the native lexicon. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1029–1040. doi: 10.1037/a0015901
- Chen, P., Bobb, S. C., Hoshino, N., and Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Res.* 1665, 50–64. doi: 10.1016/j.brainres.2017.03.023
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.* 119, 1562–1573. doi: 10.1121/1.2166600
- Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (2008). The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. *J. Acoust. Soc. Am.* 123, 414–427. doi: 10.1121/1.2804952
- Cooke, M., Lecumberri, M. L. G., Scharenborg, O., and Van Dommelen, W. A. (2010). Language-independent processing in speech perception: identification of English intervocalic consonants by speakers of eight European languages. *Speech Comm.* 52, 954–967. doi: 10.1016/j.specom.2010.04.004
- Cutler, A., Garcia Lecumberri, M. L., and Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *J. Acoust. Soc. Am.* 124, 1264–1268. doi: 10.1121/1.2946707
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *J. Acoust. Soc. Am.* 116, 3668–3678. doi: 10.1121/1.1810292
- Dijkstra, T., and Van Heuven, W. J. B. (2018). “Visual word recognition in multilinguals,” in *The Oxford Handbook of Psycholinguistics*, eds. S. Rueschemeyer and M. G. Gaskell (United Kingdom: Oxford University Press), 118–143.
- Ezzatian, P., Avivi, M., and Schneider, B. A. (2010). Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? *Speech Comm.* 52, 919–929. doi: 10.1016/j.specom.2010.04.001
- Ferré, P., and Brysbaert, M. (2017). Can Lextale-Esp discriminate between groups of highly proficient Catalan–Spanish bilinguals with different language dominances? *Behav. Res. Methods* 49, 717–723. doi: 10.3758/s13428-016-0728-y
- Festen, J. M., and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.* 88, 1725–1736. doi: 10.1121/1.400247
- Flege, J. E., and Bohn, O. S. (2021). “The revised speech learning model (SLM-r),” in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed. R. Wayland (United States: University of Cambridge), 3–83.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.* 115, 2246–2256. doi: 10.1121/1.1689343
- Fricke, M., Kroll, J. F., and Dussias, P. E. (2016). Phonetic variation in bilingual speech: a lens for studying the production–comprehension link. *J. Mem. Lang.* 89, 110–137. doi: 10.1016/j.jml.2015.10.001
- Fricke, M., Zirnstein, M., Navarro-Torres, C., and Kroll, J. F. (2019). Bilingualism reveals fundamental variation in language processing. *Biling. Lang. Cogn.* 22, 200–207. doi: 10.1017/S1366728918000482
- Garcia Lecumberri, M. L., Cooke, M., and Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Comm.* 52, 864–886. doi: 10.1016/j.specom.2010.08.014
- Golestani, N., Rosen, S., and Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: the contribution of semantics. *Biling. Lang. Cogn.* 12, 385–392. doi: 10.1017/S1366728909990150
- Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001
- Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., and Rayner, K. (2011a). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *J. Exp. Psychol. Gen.* 140, 186–209. doi: 10.1037/a0022256
- Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., and Cera, C. M. (2011b). Self-ratings of spoken language dominance: a multilingual naming test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Biling. Lang. Cogn.* 15, 594–615. doi: 10.1017/S1366728911000332
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Biling. Lang. Cogn.* 1, 67–81. doi: 10.1017/S1366728998000133
- Green, D. W., and Wei, L. (2014). A control process model of code-switching. *Lang. Cogn. Neurosci.* 29, 499–511. doi: 10.1080/23273798.2014.882515
- Guediche, S., Baart, M., and Samuel, A. G. (2020). Semantic priming effects can be modulated by crosslinguistic interactions during second-language auditory word recognition. *Biling. Lang. Cogn.* 23, 1082–1092. doi: 10.1017/S1366728920000164
- Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A., Giroud, N., Coulter, K., et al. (2021). Bilingual language experience as a multidimensional spectrum: associations with objective and subjective language proficiency. *Appl. Psycholinguist.* 42, 245–278. doi: 10.1017/S0142176420000521
- Hazan, V., and Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Lang. Speech* 43, 273–294. doi: 10.1177/00238309000430030301
- Imai, S., Walley, A. C., and Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *J. Acoust. Soc. Am.* 117, 896–907. doi: 10.1121/1.1823291
- Izura, C., Cuetos, F., and Brysbaert, M. (2014). Lextale-Esp: a test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica* 35, 49–66.
- Jegerski, J. (2018). Sentence processing in Spanish as a heritage language: A self-paced Reading study of relative clause attachment. *Lang. Learn.* 68, 598–634. doi: 10.1111/lang.12289
- Jegerski, J., and Sekerina, I. A. (2020). The processing of input with differential object marking by heritage Spanish speakers. *Biling. Lang. Cogn.* 23, 274–282. doi: 10.1017/S1366728919000087
- Ju, M., and Luce, P. A. (2004). Falling on sensitive ears: constraints on bilingual lexical activation. *Psychol. Sci.* 15, 314–318. doi: 10.1111/j.0956-7976.2004.00675.x
- Kilman, L., Zekveld, A., Hällgren, M., and Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Front. Psychol.* 5:651. doi: 10.3389/fpsyg.2014.00651
- Kimball, A. E., Shantz, K., Eager, C., and Roy, J. (2019). Confronting quasi-separation in logistic mixed effects for linguistic data: a Bayesian approach. *J. Quant. Linguist.* 26, 231–255. doi: 10.1080/09296174.2018.1499457
- Krizman, J., Bradlow, A. R., Lam, S. S. Y., and Kraus, N. (2017). How bilinguals listen in noise: linguistic and non-linguistic factors. *Biling. Lang. Cogn.* 20, 834–843. doi: 10.1017/S1366728916000444
- Kroll, J. F., and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *J. Cogn. Psychol.* 25, 497–514. doi: 10.1080/20445911.2013.799170

- Kroll, J. F., Bobb, S. C., and Hoshino, N. (2014). Two languages in mind: bilingualism as a tool to investigate language, cognition, and the brain. *Curr. Dir. Psychol. Sci.* 23, 159–163. doi: 10.1177/0963721414528511
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Lagrou, E., Hartsuiker, R. J., and Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 952–965. doi: 10.1037/a0023217
- Lemhöfer, K., and Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343. doi: 10.3758/s13428-011-0146-0
- Lenth, R. V. (2021). Emmeans: estimated marginal means, aka least-squares means. R package version 1.6.3. Available at: <https://CRAN.R-project.org/package=emmeans>
- Lüdtke, D. (2018). Ggeffects: tidy data frames of marginal effects from regression models. *J. Open Source Soft.* 3:772. doi: 10.21105/joss.00772
- Lüdtke, D. (2021). sjPlot: data visualization for statistics in social science. R package version 2.9. Available at: <https://CRAN.R-project.org/package=sjPlot>
- Luk, G., and Bialystok, E. (2013). Bilingualism is not a categorical variable: interaction between language proficiency and usage. *J. Cogn. Psychol.* 25, 605–621. doi: 10.1080/20445911.2013.795574
- MacKay, I. R., Meador, D., and Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica* 58, 103–125. doi: 10.1159/000028490
- Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230. doi: 10.1371/journal.pone.0043230
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- Marian, V., and Spivey, M. (2003). Competing activation in bilingual language processing: within-and between-language competition. *Biling. Lang. Cogn.* 6, 97–115. doi: 10.1017/S1366728903001068
- Mattys, S. L., Barden, K., and Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychon. Bull. Rev.* 21, 748–754. doi: 10.3758/s13423-013-0544-7
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- Mayo, L. H., Florentine, M., and Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *J. Speech Lang. Hear. Res.* 40, 686–693. doi: 10.1044/jslhr.4003.686
- McDonald, M., and Kaushanskaya, M. (2020). Factors modulating cross-linguistic co-activation in bilinguals. *J. Phon.* 81:100981. doi: 10.1016/j.wocn.2020.100981
- McQueen, J. M., and Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *J. Acoust. Soc. Am.* 131, 509–517. doi: 10.1121/1.3664087
- Meador, D., Flege, J. E., and MacKay, I. R. (2000). Factors affecting the recognition of words in a second language. *Biling. Lang. Cogn.* 3, 55–67. doi: 10.1017/S1366728900000134
- Mercier, J., Pivneva, I., and Titone, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Biling. Lang. Cogn.* 17, 89–117. doi: 10.1017/S1366728913000084
- Misra, M., Guo, T., Bobb, S. C., and Kroll, J. F. (2012). When bilinguals choose a single word to speak: electrophysiological evidence for inhibition of the native language. *J. Mem. Lang.* 67, 224–237. doi: 10.1016/j.jml.2012.05.001
- Montrul, S. (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. *Int. J. Biling.* 13, 239–269. doi: 10.1177/1367006909339816
- Montrul, S., and Bowles, M. (2009). Back to basics: differential object marking under incomplete acquisition in Spanish heritage speakers. *Biling. Lang. Cogn.* 12, 363–383. doi: 10.1017/S1366728909990071
- Montrul, S., Foote, R., and Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: the effects of age and context of acquisition. *Lang. Learn.* 58, 503–553. doi: 10.1111/j.1467-9922.2008.00449.x
- Morales, J., Gómez-Ariza, C. J., and Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *J. Cogn. Psychol.* 25, 531–546. doi: 10.1080/20445911.2013.807812
- Morini, G., and Newman, R. S. (2020). Monolingual and bilingual word recognition and word learning in background noise. *Lang. Speech* 63, 381–403. doi: 10.1177/0023830919846158
- Peelle, J. E. (2018). Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000312
- Peters, R. W., Moore, B. C., and Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J. Acoust. Soc. Am.* 103, 577–587. doi: 10.1121/1.421128
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312
- Quené, H., and Van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Comm.* 52, 911–918. doi: 10.1016/j.specom.2010.03.005
- R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl. Psycholinguist.* 27, 465–485. doi: 10.1017/S014271640606036X
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: romance languages as heritage languages. *Int. J. Biling.* 13, 155–163. doi: 10.1177/1367006909339814
- Scharenborg, O., Coumans, J. M., and Van Hout, R. (2018). The effect of background noise on the word activation process in nonnative spoken-word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 233–249. doi: 10.1037/xlm0000441
- Scharenborg, O., and van Os, M. (2019). Why listening in background noise is harder in a non-native language than in a native language: a review. *Speech Comm.* 108, 53–64. doi: 10.1016/j.specom.2019.03.001
- Schulpen, B., Dijkstra, T., Schriefers, H. J., and Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 1155–1178. doi: 10.1037/0096-1523.29.6.1155
- Shook, A., Goldrick, M., Engstler, C., and Marian, V. (2015). Bilinguals show weaker lexical access during spoken sentence comprehension. *J. Psycholinguist. Res.* 44, 789–802. doi: 10.1007/s10936-014-9322-6
- Skoe, E., and Karayanidi, K. (2019). Bilingualism and speech understanding in noise: auditory and linguistic factors. *J. Am. Acad. Audiol.* 30, 115–130. doi: 10.3766/jaaa.17082
- Spivey, M. J., and Marian, V. (1999). Cross talk between native and second languages: partial activation of an irrelevant lexicon. *Psychol. Sci.* 10, 281–284. doi: 10.1111/1467-9280.00151
- Syndicate, U.C.L.E. (2001). *Quick Placement Test*. Oxford: Oxford University Press.
- Tabri, D., Abou Chacra, K. M. S., and Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *Int. J. Lang. Comm. Disorder.* 46, 411–422. doi: 10.3109/13682822.2010.519372
- Takata, Y., and Nábelek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *J. Acoust. Soc. Am.* 88, 663–666. doi: 10.1121/1.399769
- Tomoschuk, B., Ferreira, V. S., and Gollan, T. H. (2019). When a seven is not a seven: self-ratings of bilingual language proficiency differ between and within language populations. *Biling. Lang. Cogn.* 22, 516–536. doi: 10.1017/S1366728918000421
- U.S. Census Bureau (2019). Language spoken at home, 2015–2019 American Community Survey 5-year estimates. Available at: <https://data.census.gov/cedsci/table?q=Language%20Spoken%20at%20Home&tid=ACST5Y2019.S1601> (Accessed June 20, 2021).
- Valdés, G. (1995). The teaching of minority languages as academic subjects: pedagogical and theoretical challenges. *Mod. Lang. J.* 79, 299–328. doi: 10.1111/j.1540-4781.1995.tb01106.x

- Valdés, G. (2000). *Spanish for Native Speakers: AATSP Professional Development Series Handbook for Teachers K-16. Vol. 1*. New York: Harcourt College.
- Van Engen, K. J. (2010). Similarity and familiarity: second language sentence recognition in first-and second-language multi-talker babble. *Speech Comm.* 52, 943–953. doi: 10.1016/j.specom.2010.05.002
- Van Engen, K. J., and Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J. Acoust. Soc. Am.* 121, 519–526. doi: 10.1121/1.2400666
- Van Hell, J. G., and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychon. Bull. Rev.* 9, 780–789. doi: 10.3758/BF03196335
- von Hapsburg, D., and Bahng, J. (2009). Effects of noise on bilingual listeners' first language (L1) speech perception. *Perspect. Hear. Hear. Disorder.* 13, 21–26. doi: 10.1044/hhd13.1.21
- Vu, V. Q. (2011). ggbiplot: A ggplot2 based biplot. R package version 0.55. Available at: <http://github.com/vqv/ggbiplot>
- Warzybok, A., Brand, T., Wagener, K. C., and Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *Int. J. Audiol.* 54, 88–99. doi: 10.3109/14992027.2015.1063715
- Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/S0749-596X(03)00105-0
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Woods, K. J., Siegel, M. H., Traer, J., and McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Atten. Percept. Psychophys.* 79, 2064–2072. doi: 10.3758/s13414-017-1361-2
- Woutersen, M., de Bot, K., and Weltens, B. (1995). The bilingual lexicon: modality effects in processing. *J. Psycholinguist. Res.* 24, 289–298. doi: 10.1007/BF02145058
- Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Fricke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [frontiersin.org/about/contact](http://frontiersin.org/about/contact)



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership