

# PROCESSING ACROSS LANGUAGES

EDITED BY : Shelia Kennison  
PUBLISHED IN : Frontiers in Psychology



# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-419-8

DOI 10.3389/978-2-88945-419-8

## 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: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# PROCESSING ACROSS LANGUAGES

Topic Editor:

**Shelia Kennison**, Oklahoma State University, United States

**Citation:** Kennison, S., ed. (2018). Processing Across Languages. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88945-419-8

# Table of Contents

- 05** ***Locality and Word Order in Active Dependency Formation in Bangla***  
Dustin A. Chacón, Mashrur Imtiaz, Shirsho Dasgupta, Sikder M. Murshed, Mina Dan and Colin Phillips
- 23** ***Experience-Based Probabilities Modulate Expectations in a Gender-Coded Artificial Language***  
Anton Öttl and Dawn M. Behne
- 37** ***How Orthography Modulates Morphological Priming: Subliminal Kanji Activation in Japanese***  
Yoko Nakano, Yu Ikemoto, Gunnar Jacob and Harald Clahsen
- 47** ***Metaphoric Reference: An Eye Movement Analysis of Spanish–English and English–Spanish Bilingual Readers***  
Roberto R. Heredia and Anna B. Cieslicka
- 57** ***Effects of Valence and Origin of Emotions in Word Processing Evidenced by Event Related Potential Correlates in a Lexical Decision Task***  
Kamil K. Imbir, Tomasz Spustek and Jarosław Żygierewicz
- 71** ***Processing Code-Switching in Algerian Bilinguals: Effects of Language Use and Semantic Expectancy***  
Souad Kheder and Edith Kaan
- 87** ***Gauging the Impact of Gender Grammaticization in Different Languages: Application of a Linguistic-Visual Paradigm***  
Sayaka Sato, Pascal M. Gyax and Ute Gabriel
- 100** ***Effects of Word Width and Word Length on Optimal Character Size for Reading of Horizontally Scrolling Japanese Words***  
Wataru Teramoto, Takuyuki Nakazaki, Kaoru Sekiyama and Shuji Mori
- 108** ***Electrophysiological Correlates of Second-Language Syntactic Processes Are Related to Native and Second Language Distance Regardless of Age of Acquisition***  
Begoña Díaz, Kepa Erdocia, Robert F. de Menezes, Jutta L. Mueller, Núria Sebastián-Gallés and Itziar Laka
- 121** ***Verbal Semantics Drives Early Anticipatory Eye Movements during the Comprehension of Verb-Initial Sentences***  
Sebastian Sauppe
- 134** ***Structural Priming and Frequency Effects Interact in Chinese Sentence Comprehension***  
Hang Wei, Yanping Dong, Julie E. Boland and Fang Yuan

- 143** *A Neurophysiological Investigation of Non-native Phoneme Perception by Dutch and German Listeners*  
Heidrun Bien, Adriana Hanulíková, Andrea Weber and Pienie Zwitserlood
- 153** *Processing Preference Toward Object-Extracted Relative Clauses in Mandarin Chinese by L1 and L2 Speakers: An Eye-Tracking Study*  
Yao-Ting Sung, Jung-Yueh Tu, Jih-Ho Cha and Ming-Da Wu
- 167** *Early Literacy Experiences Constrain L1 and L2 Reading Procedures*  
Adeetee Bhide
- 183** *Polish Pseudo-Words List: Dataset of 3023 Stimuli with Competent Judges' Ratings*  
Kamil K. Imbir, Tomasz Spustek and Jarosław Żygierewicz
- 186** *Co-Lateralized Bilingual Mechanisms for Reading in Single and Dual Language Contexts: Evidence from Visual Half-Field Processing of Action Words in Proficient Bilinguals*  
Marlena Krefta, Bartosz Michałowski, Jacek Kowalczyk and Gregory Króliczak



# Locality and Word Order in Active Dependency Formation in Bangla

Dustin A. Chacón<sup>1,2\*</sup>, Mashrur Imtiaz<sup>3</sup>, Shirsho Dasgupta<sup>4</sup>, Sikder M. Murshed<sup>3</sup>, Mina Dan<sup>4</sup> and Colin Phillips<sup>1</sup>

<sup>1</sup> Department of Linguistics, University of Maryland, College Park, College Park, MD, USA, <sup>2</sup> Department of Linguistics, University of Minnesota, Minneapolis, MN, USA, <sup>3</sup> Department of Linguistics, University of Dhaka, Dhaka, Bangladesh, <sup>4</sup> Department of Linguistics, University of Calcutta, Kolkata, India

Research on filler-gap dependencies has revealed that there are constraints on possible gap sites, and that real-time sentence processing is sensitive to these constraints. This work has shown that comprehenders have preferences for potential gap sites, and immediately detect when these preferences are not met. However, neither the mechanisms that select preferred gap sites nor the mechanisms used to detect whether these preferences are met are well-understood. In this paper, we report on three experiments in Bangla, a language in which gaps may occur in either a pre-verbal embedded clause or a post-verbal embedded clause. This word order variation allows us to manipulate whether the first gap linearly available is contained in the same clause as the filler, which allows us to dissociate *structural locality* from *linear locality*. In Experiment 1, an untimed ambiguity resolution task, we found a global bias to resolve a filler-gap dependency with the first gap linearly available, regardless of structural hierarchy. In Experiments 2 and 3, which use the filled-gap paradigm, we found sensitivity to disruption only when the blocked gap site is both structurally and linearly local, i.e., the filler and the gap site are contained in the same clause. This suggests that comprehenders may not show sensitivity to the disruption of all preferred gap resolutions.

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University–Stillwater,  
USA

### Reviewed by:

Robert Frank,  
Yale University, USA  
Ina Bornkessel-Schlesewsky,  
University of South Australia, Australia

### \*Correspondence:

Dustin A. Chacón  
dustin@umn.edu

**Keywords:** filler-gap dependencies, locality, Bangla, sentence processing, islands

## INTRODUCTION

The formation of linguistic dependencies is subject to a wide variety of constraints. Some constraints are conditions on grammatical well-formedness, whereas others define the interpretations that are preferred in real-time sentence processing. Locality constraints on filler-gap dependencies are one particularly well-studied example of both constraint types. Some locality constraints distinguish acceptable filler-gap dependencies from unacceptable filler-gap dependencies, as long recognized by syntacticians (Ross, 1967; Huang, 1982; Rizzi, 1982, 1990, 2013; Chomsky, 1986; Rudin, 1988; Lasnik and Saito, 1992; Manzini, 1992; Szabolcsi and den Dikken, 1999; Boeckx, 2008). For instance, the filler-gap dependency in (1a) between *who* and the position in which it is interpreted (marked as     ) is judged acceptable, in contrast with the sentence in (1b). This is because filler-gap dependencies may not cross into clauses (marked *S'*) in the subject position of another clause (this violates the *sentential subject constraint* and the *complex noun phrase constraint*, Ross, 1967). Constraints on acceptable filler-gap dependencies are called *island constraints*.

### Specialty section:

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

**Received:** 02 December 2015

**Accepted:** 03 August 2016

**Published:** 25 August 2016

### Citation:

Chacón DA, Imtiaz M, Dasgupta S,  
Murshed SM, Dan M and Phillips C  
(2016) Locality and Word Order in  
Active Dependency Formation in  
Bangla. *Front. Psychol.* 7:1235.  
doi: 10.3389/fpsyg.2016.01235

- (1) a. I know **who** it surprised Dale [<sub>S</sub> that Sarah saw \_\_\_\_].  
 b. \*I know **who** [<sub>S</sub> that Sarah saw \_\_\_\_] surprised Dale.

Other locality constraints determine which gap sites are preferred when multiple possibilities are available. In on-line tasks, this manifests as a preference for early resolution, a process called *active dependency formation* (Fodor, 1978; Crain and Fodor, 1985; Stowe, 1986; Frazier, 1987; Frazier and Flores d'Arcais, 1989). For instance, Stowe (1986) observed longer reading times at the direct object *us* in (2a) compared to the control sentence in (2b), which lacks a filler-gap dependency. This increase in reading times, called the *filled-gap effect*, suggests that readers make an early commitment to resolve *who* as the direct object of *bring* before it is clear whether there is a direct object gap. Encountering the direct object pronoun *us* then triggers a reanalysis process, leading to an increase in processing difficulty.

- (2) a. My brother wanted to know **who** Ruth would bring us home to \_\_\_\_ at Christmas.  
 b. My brother wanted to know if Ruth would bring us home to somebody at Christmas.

There has been much interest in determining whether these two types of constraints are the same, following from some independently motivated restrictions on linguistic processes, e.g., restrictions on memory capacity (Deane, 1991; Pritchett, 1992; Kluender and Kutas, 1993; Kluender, 1998, 2004; Hofmeister and Sag, 2010; for discussion see Phillips, 2013). Explaining island phenomena as a consequence of resource limitation has the potential to radically simplify grammatical theories.

If island constraints are indeed reducible to constraints on preferred gap sites, then both sets of constraints should be sensitive to the same properties of the linguistic representation being computed. In other words, the notion of “local” that is relevant should be the same. It is relatively uncontroversial that island constraints are defined in terms of formal linguistic structure, either hierarchical syntactic relations (Ross, 1967; Chomsky, 1981, 1986; Huang, 1982; Rizzi, 1990; Lasnik and Saito, 1992; for review, see Rizzi, 2013), or semantic/pragmatic relations (Erteschik-Shir, 1973; Kuno, 1976; Szabolcsi and Zwarts, 1993; Truswell, 2007; Ambridge and Goldberg, 2008; Abrusán, 2011a,b). However, it is unclear what notion of locality is relevant for determining preferred gap sites. For instance, the direct object position of *bring* in (2b) may be preferred, because fewer nodes separate this gap site from the filler compared to other potential forthcoming gap sites, i.e., there is an additional PP node separating the filler and prepositional object gap site, illustrated in (3). To construct the direct object gap, the comprehender needs to postulate a less articulated structure (a verb phrase and an object position) than in alternative analyses (a verb phrase, plus dependents on this verb phrase, such as a prepositional phrase, and an object position). Alternatively, the direct object position may be preferred because it is the first position that is linearly available. That is, the locality constraints on preferred gap sites may be defined in terms of *structural locality* or *linear locality*. If the constraints on preferred gap sites are sensitive to linear locality, then this motivates maintaining a

distinction between island constraints and locality constraints on preferred gap sites.

- (3) My brother wanted to know  
**who** [<sub>S</sub> Ruth would [<sub>VP</sub> bring us home [<sub>PP</sub> to \_\_\_\_] at Christmas]]]

Most research on filler-gap dependency processing cannot decide among these hypotheses, because most studies are conducted on languages like English, where structural and linear locality converge, as illustrated above. However, previous work on Japanese, a language with different word order properties than English, suggests that these constraints are dissociated (Aoshima et al., 2004; Yoshida, 2006; Omaki et al., 2013). This is discussed in more detail in Section Locality in Filler-Gap Dependencies.

In this paper, we report on three experiments in Bangla (Bengali) that further investigate locality constraints on preferred gap sites. Bangla is a valuable language for this purpose, because embedded clauses may either precede or follow the embedding verb, as shown in (4) and (5). Additionally, Bangla allows filler-gap dependencies with *wh*-phrases. These filler-gap dependencies may resolve in either the main clause, or an embedded clause on either side of the main verb, as shown in (6) and (7). This allows us to manipulate whether the first gap site is structurally local or distant within the same language, which allows a within-language comparison of the influence of word order on filler-gap dependency processing, which has previously only been conducted in a cross-language fashion (Omaki et al., 2013).

- (4) raj bollo [<sub>S</sub> še ašbe \_\_\_\_ ]  
 Raj said he come.FUT  
 “Raj said that he will come.”
- (5) raj [<sub>S</sub> še ašbe \_\_\_\_ ] bollo  
 Raj he come.FUT said  
 “Raj said that he will come.”
- (6) raj **kəkhon** \_\_\_\_ bollo [<sub>S</sub> še \_\_\_\_ ašbe \_\_\_\_ ]  
 raj when said he come.FUT  
 “When did Raj say \_\_\_\_ that he will come \_\_\_\_ ?”
- (7) raj **kəkhon** [<sub>S</sub> še \_\_\_\_ ašbe \_\_\_\_ ] \_\_\_\_ bollo  
 raj when he come.FUT said  
 “When did Raj say \_\_\_\_ that he will come \_\_\_\_ ?”

Experiment 1 was a within-language replication of the cross-language findings from Omaki et al. (2013). In Experiment 1, we investigated how ambiguous filler-gap dependencies like (6) and (7) are resolved using an off-line ambiguity resolution task. This task allows us to probe for preferences directly, instead of relying on an indirect measure, such as increased reading times indicating detection of an unexpected parse. We found that filler-gap dependencies are resolved with the first position linearly available across word orders. In main verb first word orders as in (6), the filler-gap dependency was resolved with the main verb. In embedded verb first word orders like (7), it was resolved with the embedded verb.

In Experiment 2, we investigated the preference for linearly local gap sites in an on-line, filled-gap paradigm task. This task provides a more standard measure of disruption in

moment-by-moment sentence comprehension, and thus it can be used to determine the time course of active dependency formation. Like Experiment 1, we leveraged the flexible word order of Bangla to manipulate whether the first available gap site was in the same clause as the filler or in an embedded clause. We found a filled-gap effect when resolution with the first gap site was blocked in main verb first word orders like (6), where structural locality and linear locality aligned, but not in embedded verb first word orders like (7). In other words, there was only detection of a blocked filler-gap resolution when the gap site was both structurally local and linearly local, but not when this position was structurally distant. The comparison between Experiments 1 and 2 suggests a contrast between gap site preferences and sensitivity to disruption.

The apparent mismatch in Experiments 1 and 2 may be due to the on-line/off-line contrast between the two experiments, or to the ambiguity resolution/filled-gap paradigm difference. In Experiment 3, we diagnosed the cause of this mismatch. Experiment 3 was an off-line acceptability judgment task, like Experiment 1, that used the filled-gap paradigm, like Experiment 2. We again only found evidence that comprehenders detected a filled-gap when the filler-gap dependency was blocked from resolving with a structurally local and linearly local position, as in Experiment 2. This suggests that the contrast between locality preferences and sensitivity to disruption for embedded verb first word orders in Experiments 1 and 2 was not due to the off-line/on-line contrast, but rather the specific mechanisms underlying filled-gap detection.

## LOCALITY IN FILLER-GAP DEPENDENCIES

There is substantial evidence that shorter filler-gap dependencies are preferred to longer filler-gap dependencies. For instance, Frazier and Clifton (1989) found that reading times were increased for sentences containing filler-gap dependencies spanning multiple clauses compared to controls (see also Kluender and Kutas, 1993; Dickey, 1996; Kluender, 1998). This bias against longer filler-gap dependencies is also reflected in offline acceptability judgments, where sentences containing filler-gap dependencies spanning multiple clauses are rated lower than sentences with shorter filler-gap dependencies (Phillips et al., 2005; Alexopoulou and Keller, 2007; Sprouse et al., 2012).

Online studies show that the preference for shorter filler-gap dependencies manifests as a preference for early resolution. For instance, the filled-gap effect discussed in Section Introduction demonstrates that blocking an early filler-gap dependency resolution triggers a costly reanalysis process (Crain and Fodor, 1985; Stowe, 1986; Lee, 2004). Converging evidence comes from the plausibility mismatch paradigm (Garnsey et al., 1989; Traxler and Pickering, 1996). For instance, in a series of eye-tracking experiments, Traxler and Pickering (1996) observed that gaze times increased on the verb *wrote* in (8b) compared to (8a).

(8) a. We like **the book** that the author wrote unceasingly and with great dedication about \_\_\_ while waiting for a contract.

b. We like **the city** that the author wrote unceasingly and with great dedication about \_\_\_ while waiting for a contract.

This suggests that *the city* was first interpreted as the object of *wrote*. Comprehenders could then detect that the early gap commitment yields an implausible interpretation. Then, they rejected this commitment, and searched for a different gap, yielding a reanalysis cost. Thus, like we argued for the filled-gap effect, the plausibility mismatch effect illustrates not only early commitment to a local gap, but also sensitivity to disruption when this position is unavailable. Other converging evidence for active dependency formation comes from EEG studies (Garnsey et al., 1989; Kaan et al., 2000; Phillips et al., 2005), the “stops making sense” task (Tanenhaus et al., 1985; Boland et al., 1995), cross-modal lexical priming (Nicol and Swinney, 1989; Nicol et al., 1994), and “visual world” eye-tracking (Sussman and Sedivy, 2003).

This bias toward early filler-gap dependency resolution in real-time behavior and toward shorter dependencies in offline judgments is commonly attributed to resource limitations. For instance, unintegrated fillers may require memory resources to be actively maintained (Jackendoff and Culicover, 1971; Wanner and Maratsos, 1978). Alternatively, longer dependencies in general may be more costly, leading to a dispreference for longer filler-gap dependencies (Gibson, 1998; Hawkins, 2004). Other analyses contend that longer filler-gap dependencies may cause increased processing difficulty because the filler must be retrieved from memory at the gap site, which may be costly and error-prone in the case of longer dependencies (McElree, 2006; Wagers and Phillips, 2014). Lastly, more local gaps may be preferred because comprehenders attempt to resolve as many grammatical requirements as early as possible (Pritchett, 1992; Weinberg, 1992; Altmann and Kamide, 1999; Sedivy et al., 1999; Aoshima et al., 2004; Wagers and Phillips, 2009). These accounts all imply that the comprehender should minimize filler-gap dependency length in order to optimize resource usage. However, these accounts make no commitment as to whether linear locality or structural locality are relevant in selecting preferred gap sites.

Island constraints, in contrast, are typically described in structural terms. Island constraints are restrictions on possible filler-gap dependencies, with several illustrated in

- (9) a. Relative Clause Island:  
\* **Who** did Dale comfort [<sub>NP</sub> the woman that [<sub>S</sub> saw \_\_\_ ?]]
- b. Whether Island:  
\* **Who** did Dale wonder [whether Bob frightened \_\_\_ ?]
- c. *Wh*-Island:  
\* **Who** did Dale say [who saw \_\_\_ behind Laura's bed?]
- d. Subject Island:  
\* **Who** did [the fact that Sarah saw \_\_\_] surprise Dale?
- e. Adjunct Island:  
\* **Who** did Dale ruminate [while Harry interrogated \_\_\_ ?]
- f. Coordinate Structure Constraint:  
\* **Who** did [Dale suspect \_\_\_ and Harry interrogate Leland?]



## g. Factive Island:

\* **Why** did Dale remember [that Ben was suspicious \_\_\_?]

Island constraints have long been studied in theoretical linguistics, where they typically are characterized as constraints on well-formed linguistic representations, either as formal syntactic constraints (Ross, 1967; Chomsky, 1977, 1981, 1986; Huang, 1982; Rizzi, 1990, 2013; Lasnik and Saito, 1992), or as constraints on well-formed and felicitous semantic/pragmatic forms (Erteschik-Shir, 1973; Kuno, 1976; Szabolcsi and Zwarts, 1993; Truswell, 2007; Ambridge and Goldberg, 2008; Abrusán, 2011a,b). As such, island constraints are typically defined over the hierarchical structure of the sentence, or the formal relations between the words and phrases. This can be demonstrated with pairs like (10), repeated from (1), in which the filler-gap dependency that spans fewer words is dispreferred to a filler-gap dependency that spans more words. This contrast can be characterized as a formal constraint against gaps in subject clauses, but not extraposed clauses (Ross, 1967).

- (10) a. I know **who** it surprised Dale [<sub>S</sub> that Sarah saw \_\_\_] ?  
 b. \* I know **who** [<sub>S</sub> that Sarah saw \_\_\_] surprised Dale?

Island constraints are observed to be robust in both off-line and on-line measures. Off-line acceptability judgments show that speakers give low ratings to sentences with island violations (Sobin, 1987; Cowart, 1996, 2003; Alexopoulou and Keller, 2007; Heestand et al., 2011; Sprouse et al., 2012). Additionally, the effects of active dependency formation typically disappear in island constructions. There are no filled-gap effects or plausibility mismatch effects inside island contexts (Stowe, 1986; Bourdages, 1992; Traxler and Pickering, 1996). Similarly, results from EEG studies (Neville et al., 1991; Kluender and Kutas, 1993; McKinnon and Osterhout, 1996) and speed-accuracy tradeoff studies (McElree and Griffith, 1998) suggest that comprehenders immediately detect island boundaries. The rapid application of island constraints can be explained in theories of sentence processing that posit rapid and faithful use of grammatical constraints (e.g., Lewis and Phillips, 2015) or theories that posit that representations with gap sites inside island contexts are too costly to represent (Gibson, 1998; Hawkins, 2004).

Some data suggests the constraints on preferred gaps should be dissociated from island constraints (Phillips, 2006; Wagers and Phillips, 2009; Sprouse et al., 2012; Yoshida et al., 2014). Other findings imply that constraints on preferred gap sites are defined in terms of linear locality, unlike island constraints which are defined in terms of structural locality. These findings come from Japanese, a language in which embedded clauses precede the main verb, meaning that in multi-clause sentences structural positions that are linearly closer may be structurally more distant. This makes it possible to dissociate structural locality and linear locality. Japanese speakers prefer to resolve filler-gap dependencies in embedded clauses, likely because this is the first position linearly available. For instance, Aoshima et al. (2004) found filled-gap effects for sentences like (11), in which the fronted dative phrase *dono-syain-ni* “which employee-DAT” was blocked from resolving with the embedded clause because of the case-matched noun phrase *kacyoo-ni* “assistant

manager-DAT” (see also Yoshida, 2006). Similarly, Omaki et al. (2013) showed that speakers of Japanese interpreted an ambiguously fronted *wh*-phrase, as in (12), with the embedded clause in a Question after Story task, a task that provides an untimed measure of how speakers prefer to interpret ambiguous questions (de Villiers et al., 1990). This shows that in off-line measures of gap location preferences and on-line measures of filled-gap detection, Japanese speakers prefer a linearly local resolution.

- (11) **Dono-syain-ni** senmu-wa  
 which employee-DAT managing director-TOP  
 [syacyoo-ga kaigi-de  
 president-NOM meeting-at  
 kacyoo-ni syookyuu-o yakusoku-sita-to]  
 assistant manager-DAT raise-ACC promised-DECLC  
 iimasita-ka?  
 told-Q?  
 “**Which employee** did the managing director  
 tell \_\_\_ that the president promised a raise to  
 the assistant manager at the meeting?”
- (12) **Doko-de** Yukiko-chan-wa [choucho-o  
 where-at Yukiko-DIM-TOP butterfly-ACC  
 tsukumaeru-to] itteta-no?  
 catch-DECLC was telling-Q?  
 “Where did Yukiko say that she will catch butterflies?”

In this paper, we further investigate this generalization in Bangla, a language with variable word order that permits us to manipulate whether the most linearly local potential gap site is within the same clause as the fronted filler (i.e., structurally local), or in an embedded clause (i.e., structurally non-local). In Section Grammatical Properties of Bangla, we describe the relevant properties of Bangla syntax. In Sections Experiment 1–General Discussion we describe the results of three experiments on Bangla filler-gap dependency processing.

## GRAMMATICAL PROPERTIES OF BANGLA

Bangla is a language spoken primarily in Bangladesh and the eastern Indian state of West Bengal, with approximately 180 million speakers worldwide (Lewis et al., 2015). Bangla is in the Eastern Zone of the Indo-Aryan branch of the Indo-European language family. Due to its contact with multiple linguistic areas, Bangla has many properties typical of northern Indo-Aryan, Dravidian, and Southeast Asian languages. For more complete descriptions of the language, see Thompson (2010) and David (2015).

Embedded clauses in Bangla may either precede or follow an embedding verb, shown in (13). Post-verbal embedded clauses may be introduced with the complementizer *je*, shown in (14a). Pre-verbal embedded clauses may appear with the complementizer *bole* at the end of the clause, shown in (14b), or with *je* in a clause-internal position, shown in (14c). Dasgupta (2007) describes the clause-internal *je* as an “anchor,” which may

be a distinct lexical category. Examples are taken from Bayer (1996).

- (13) a. *še bollo ora ašbe*  
he said they come.FUT
- b. *še ora ašbe bollo*  
he they come.FUT said  
'He said that they will come'
- (14) a. *chele-ṭa bollo [s' je tar baba ašbe ]*  
boy-CL said that his father come.FUT
- b. *chele-ṭa [s' tar baba ašbe bole ] bollo*  
boy- CL his father come.FUT that said
- c. *chele-ṭa [s' tar baba je ašbe ] bollo*  
boy- CL his father that come.FUT said  
'The boy said that his father will come'

These constructions are used in similar contexts, although there are subtle syntactic and semantic differences that we leave aside (for discussion see Bal, 1990 on related constructions in Oriya, and Bayer, 1996, 1999, 2001; Simpson and Bhattacharya, 2000, 2003).

Case-marking is often an important cue in detecting clause boundaries in head-final languages. For example, Japanese speakers use nominative-marked noun phrases to detect the beginning of embedded clauses (Miyamoto, 2002). We assume that Bangla speakers do the same, although we have not directly tested this. Bangla has four cases—nominative, accusative, genitive, and oblique. The first three cases are clearly marked in the pronoun system, e.g., *še* "3SG.NOM," *take* "3SG.ACC," and *tar* "3SG.GEN." Thus, in (13b), the comprehender can detect the embedded clause, because *ora* "3PL.NOM" is a clearly nominative-marked pronoun, as is *še* "3SG.NOM." For other noun phrases, nominative case is left unmarked, and the accusative case morpheme (*-ke*) is reserved for animate objects or specific inanimate objects. In (14b–14c), a comprehender can detect the embedded clause at *baba*, "father." This is because *baba* "father" is an animate noun that is not marked with an overt accusative, genitive, or oblique morpheme. Thus, it must be nominative. Given that there was a previous nominative noun phrase (*chele-ṭa* "the boy"), the comprehender should postulate an embedded clause here, as well.

Like English and Japanese, Bangla also permits unbounded filler-gap dependencies. Gaps may either occur in pre-verbal or post-verbal embedded clauses. Extraction from a post-verbal clause is shown in (15), adapted from Simpson and Bhattacharya (2003). In (15a), the noun phrase *hæmleṭ* "Hamlet" is interpreted as the direct object of the verb *poṛeche* "read." In (15b) and (15c), *hæmleṭ* "Hamlet" appears either one or two clauses away from the embedded clause, but is still interpreted as the direct object of *poṛeche* "read." The filler may appear either after the subject or before the subject, as in (15d).

- (15) a. *jɔn bhablo [s' meri bollo [s' su hæmleṭ poṛeche ]]*  
John thought Mary said Sue Hamlet read

- b. *jɔn bhablo [s' meri hæmleṭ bollo [s' su \_\_\_\_\_*  
John thought Mary Hamlet said Sue  
*poṛeche ]]*  
read
- c. *jɔn hæmleṭ bhablo [s' meri bollo [s' su \_\_\_\_\_*  
John Hamlet thought Mary said Sue  
*poṛeche ]]*  
read  
'John thought that Mary said that Sue has read Hamlet'
- d. *hæmleṭ jɔn bhablo [s' meri bollo [s' su \_\_\_\_\_*  
Hamlet John thought Mary said Sue  
*poṛeche ]]*  
read  
'John thought that Mary said that Sue has read Hamlet'

Extraction from pre-verbal clauses is shown in (16). In (16a), the noun phrase *tomar beṛal-ke* "your cat-ACC" is interpreted as the object of the embedded verb *kamṛeche* "bit," but it appears in the left edge position of the main clause. Similarly, in (16b), the prepositional phrase *bas theke* "bus from" appears in the left edge position of the main clause, but is interpreted as a modifier of the embedded clause. This contrasts with other languages with both pre-verbal and post-verbal clauses, like Basque which disallows gap sites in pre-verbal clauses (Uriagereka, 1992), and Malayalam which only allows direct object gaps in pre-verbal clauses, but not for adjunct phrases like *bas theke* "bus from" (Srikumar, 2007). The filler may again either appear before the subject or after the subject, as in (16c).

- (16) a. **tomar beṛal-ke** amra šobai  
your cat-ACC we everyone  
*[s' paš-er baṛi-r kukur \_\_\_\_\_ kamṛeche bole ]*  
neighbor-GEN dog bit that  
*šunechilam*  
heard  
'We had all heard that the neighbor's dog has bitten your cat'
- b. **bas theke** amar didi  
bus from my sister  
*[s' ɔtogulo duronto bacca laphiye*  
so many uncontrollable child jumping  
*nambe bole ] bhabbe ni*  
descend.FUT that think PST.NEG  
'My sister hasn't thought that so many children could jump down from a bus.'
- c. **amar didi bas theke**  
my sister bus from  
*[s' ɔtogulo duronto bacca laphiye*  
so many uncontrollable child jumping  
*nambe bole ] bhabbe ni*  
descend.FUT that think PST.NEG

‘My sister hasn’t thought that so many children could jump down from a bus.

To summarize, Bangla permits embedded clauses to precede or follow the embedding verb. Additionally, fillers in the main clause may resolve with gap sites in the main clause or in an embedded clause on either side of the embedding verb. This means the schematic representations in (17) are all permissible, making Bangla an excellent language for testing locality biases.

- (17) a. Post-verbal embedded clause, main clause resolution:  
 ... *filler* ... \_\_\_\_ ... V ... [*S'*... ] ...  
 b. Post-verbal embedded clause, embedded clause resolution:  
 ... *filler* ... V ... [*S'*... \_\_\_\_ ] ...  
 c. Pre-verbal embedded clause, embedded clause resolution:  
 ... *filler* ... [*S'*... \_\_\_\_ ] ... V ...  
 d. Pre-verbal embedded clause, main clause resolution:  
 ... *filler* ... [*S'*... ] ... \_\_\_\_ ...V ...

If the locality constraints on preferred gap sites are sensitive to linear order, as suggested by findings in Japanese, then the dependencies schematized in (17a) and (17c) should be preferred to those in (17b) and (17d). However, if locality constraints on preferred gap sites are sensitive to structural locality, then the representations in (17a) and (17d) should be preferred, since the filler and gap site are structurally more local to the filler. We test these predictions in Experiments 1–3.

## EXPERIMENT 1

### Rationale

In Experiment 1, we used the Question after Story task (de Villiers et al., 1990) to determine whether Bangla speakers prefer linearly local gap sites across word orders. We adapted the design used by Omaki et al. (2013), which probed for word order effects on filler-gap dependency resolution using a between language comparison. In their study, participants viewed a series of vignettes in which a character acted out an event in one location and reported on it in another location. Afterwards, participants were asked to respond to a question that contained a fronted *wh*-filler that could resolve in either the embedded clause or main clause. Participants’ responses revealed in which clause they preferred to resolve the filler-gap dependency. In English, a language that conflates linear and structural locality, the ambiguous filler-gap dependency was most commonly resolved with the main clause in Omaki and colleagues’ studies. Conversely, in Japanese, the filler-gap dependency was preferentially resolved in the embedded clause. They took this as evidence for a universal preference to resolve filler-gap dependencies with the first position linearly available.

Our study took advantage of the flexible word order in Bangla to further test this hypothesis. The study had two main conditions: a main verb first condition, shown in (18a), and an embedded verb first condition, shown in (18b). For both sentences, the fronted *wh*-filler *kothae* “where” could be resolved in the embedded clause, modifying the catching event, or the

main clause, modifying the telling event. If gaps are preferentially constructed in the first position linearly available, as suggested by Omaki and colleagues’ cross-language contrast, then we expected *kothae* “where” to be resolved with the main verb in word orders like (18a), and with the embedded verb in word orders like (18b).

### (18) a. Main Verb First Condition:

šumi kothae ækjɔn-ke boleche [*S'* je še  
 Shumi where someone-ACC told that she  
 prɔjapoti dhorbe]?  
 butterfly catch.FUT

### b. Embedded Verb First Condition:

šumi kothae [*S'* še prɔjapoti dhorbe bole]  
 Shumi where she butterfly catch.FUT that  
 ækjɔn-ke boleche?  
 someone-ACC told  
 “Where did Shumi tell someone that she will catch  
 butterflies?”

## Participants

Ninety-six participants were recruited for Experiment 1. Forty-eight adult native speakers of Bangla were collected from the student population at The University of Dhaka in Dhaka, Bangladesh, and 48 participants were from the student population at Calcutta University in Kolkata, India. Bangladeshi participants were compensated 500 Bangladeshi Taka (BDT), and Indian participants were compensated 200 Indian Rupees (INR). This session took approximately 15 min. Experiment 1 was conducted after participants completed either Experiment 2 or after another experiment unrelated to the current study. These populations were each split into two groups, a “within-subjects” and a “between-subjects” group, as discussed in section Materials. We tested participants in both India and Bangladesh to probe for any potential influence of dialect difference, especially given that Indian Bangla speakers are likely to be competent in Hindi, which uses different *wh*-scope marking strategies (e.g., Dayal, 1996; Manetta, 2012). Additionally, we included a within-subjects and between-subjects manipulation to check for any effect of self-priming in the experiment. This was important for comparing our within-language findings to results from previous between-language comparisons, where participants in each language, e.g., Japanese and English, saw only one of the word orders tested in Bangla.

## Materials

The materials were adapted from Omaki et al. (2013). The stories and audio were translated by three of the authors to standard colloquial Dhakaia Bangla. Some lexical material was changed to better suit the different cultural context, including names. The questions were presented on a paper questionnaire. Participants were instructed to respond to a question printed on the questionnaire immediately after each vignette, before progressing onto the next vignette. Across all questionnaires, we rigidly alternated between a target item and a filler item, in order to reduce priming or perseveration effects. The target items

were two-clause sentences with an ambiguous *wh*-dependency, presented in (18). The fillers were one-clause sentences with an unambiguous *kæno* “why” question.

Participants were split into two groups—the “between participants” group and the “within participants” group. The “between participants” group was included to make a closer comparison to the existing literature comparing English and Japanese. The division of participants is illustrated in **Table 1**. Questionnaires were prepared for each group. For the “between participants” questionnaires, the target items all had either main verb first word orders or embedded verb first word orders, i.e., participants saw 4 target items in one of the two conditions. The remaining participants received a “within participants” questionnaire, where the target items contained both verb first word order and embedded verb first word orders, i.e., 2 target items per condition. In the within participants questionnaire, the two conditions alternated, such that there were two questions of each word order in each questionnaire.

The stories were animated vignettes made from clipart images. In each vignette, a character went to four different locations, and performed an action in each. A sample story from the English study in Omaki et al. (2013) is presented in (19). The videos are included as Supplementary Material.

(19) Sample story:

**[Introduction]**

It was a beautiful day in spring so Lizzie decided she was going to go catch butterflies in the park.

**[1st Location]**

Her Mom and Dad weren’t home, so Lizzie thought she should tell her brother or sister about going to the park, so that Mom and Dad would know where she was when they got back. She first went to her brother’s room, but he was taking a nap and she couldn’t tell him about catching butterflies.

**[2nd Location]**

Instead, Lizzie looked for her sister. She looked all over the house but didn’t see her sister anywhere! When she was about to give up, Lizzie heard her sister’s voice in the

basement! She went to the basement and said to her sister: “I’m gonna catch butterflies in the park!”

**[3rd Location]**

Then, on her way to the park, Lizzie passed by a parking lot and saw a butterfly near it. She walked slowly toward the butterfly, but before Lizzie could get there, another girl came along and caught the butterfly! Lizzie didn’t see any more butterflies there, so she kept walking toward the park.

**[4th Location]**

There were lots and lots of butterflies in the park, and she caught one in a jar and took it home with her. She liked the one that she caught, but she wished she could have caught more butterflies.

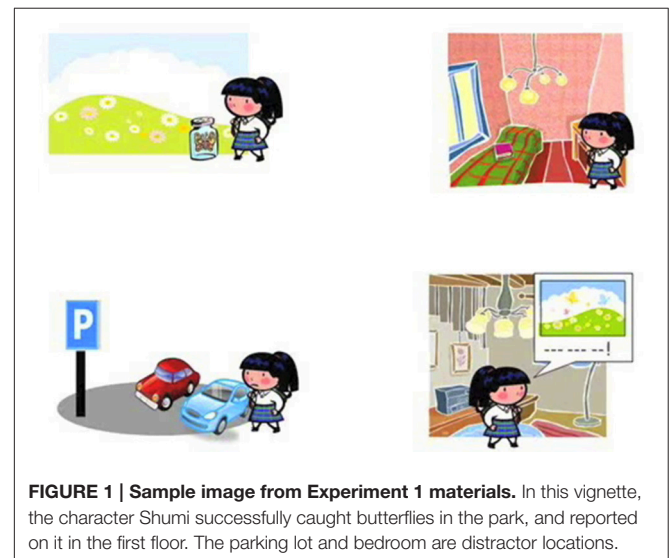
Each vignette consisted of six phases. The first phase introduced the protagonist, displayed in the center of the screen. The following four phases depicted him or her at each of the four locations. The protagonist succeeded or failed to perform some intended action as announced in the introductory phase, or succeeded or failed to report on it. The contrast between successes and failures was intended to make the event-location pairings more memorable, and to ensure that the “where” test questions were felicitous. In locations where the protagonist succeeded on performing his or her stated action or reported on it, there was a visual trace left behind (i.e., a butterfly in a bottle, or a word balloon). The first two and last two locations were relevant for either the main clause event (i.e., the reporting event), or the embedded clause event (i.e., the intended action). In the sixth and final phase, the protagonist returned to the center of the screen, and then the story concluded. A sample image from the vignette is given in **Figure 1**.

To avoid any potential recency bias, the ordering of the events within each story was counterbalanced, such that the first pair of events pertained to the reporting event in half of the stories, and to the embedded clause event in the other half of the stories. In each case, the story provided motivation for continuing to the next series of events. For instance, In (19), the

**TABLE 1 | Distribution of participants in experiment 1.**

Total: 96				
Dhaka: 48		Kolkata: 48		
Within participants: 24	Between participants: 24	Within participants: 24	Between participants: 24	
	Main verb first: 12	Embedded verb first: 12	Main verb first: 12	Embedded verb first: 12

*There were 96 participants in Experiment 1, 48 for each city, Dhaka and Kolkata. Each city was split into two groups. One group of 24 in each city saw both conditions in the same questionnaire, the within-participants group. Another group of 24 in each city was further divided in two groups of 12, one seeing only lists with main verb first word orders and the other seeing only lists with embedded verb first word order.*



**FIGURE 1 | Sample image from Experiment 1 materials.** In this vignette, the character Shumi successfully caught butterflies in the park, and reported on it in the first floor. The parking lot and bedroom are distractor locations.

reporting events are motivated by the character's need to tell her siblings where she was going. The pairings of quadrant position and event were randomized across stories so that participants could not predict which locations would correspond to which actions.

## Methods

Experiment 1 was an adaptation of Omaki et al. (2013), question-after-story task (de Villiers et al., 1990). Participants were instructed in Bangla to watch a sequence of 8 vignettes. At the end of each vignette, the screen displayed "write your answer now" in Bangla. At this point, the experimenter paused the video and instructed the participant to read a question printed on a paper questionnaire. Participants were instructed to write a brief response. We asked that the responses be brief because in pilot studies, participants attempted to recapitulate large portions of the story, which complicated coding the results. After responding, the experimenter resumed the video, which progressed to the next vignette.

## Results

We coded each response as either a main clause response or an embedded clause response, depending on which location the participant named. Responses that either failed to answer the question or that provided both possible answers were excluded. Most of the excluded responses named both possible locations, implying that Bangla speakers were often aware of the ambiguity. The proportions of excluded observations are given in Table 2.

There were fewer exclusions for the embedded verb first conditions in the between-participants conditions compared to other conditions. This is the only list in which participants saw only the canonical, verb-final word order. This is because the fillers across all lists used this word order, and all target items in this list also used embedded verb first word order. The presence of non-canonical word orders in other lists may have made the ambiguity more salient, leading to a higher number of exclusions. After excluding these observations, participants responded with the main verb location in 81% of the main verb first word orders, but only 23% of the embedded verb first word orders.

Using the lmer package in R (Bates et al., 2015), we submitted the results to a logit mixed effects model with a *bobyqa* optimizer. The predicted variable was main clause response, coded as 1. For fixed effects, we included word order (main verb first or embedded verb first), location (Dhaka or Kolkata), and list type (within participants or between participants), with their interaction terms. We included these factors in order

to fit a maximal model that tested for all potential variables of interest. For random effects, we included participant and items. Afterwards, we used the backward elimination method to eliminate factors from the model one-by-one to minimize the AIC (Akaike Information Criterion) of the model, as described by Faraway (2002). The results of the best-fit model are given in Table 3. The *p*-values in Table 3 were generated using the lmerTest package (Kuznetsova et al., 2015). The mean proportion of main verb responses is actually given in Figure 2.

We found a significant effect of word order on the proportion of main clause responses. The effect was as predicted: for the main verb first word order, participants showed a strong bias to answer with main verb locations. With embedded verb first word orders, there was a strong bias to answer with embedded verb locations. There was no significant effect of city, implying that there were no systematic dialect differences detected in Experiment 1. Additionally, there was no significant effect of list type, i.e., participants typically responded with the event denoted by the first verb linearly available regardless of whether they saw lists with only one word order or lists with mixed word order. However, there was a marginal interaction of city and list type, due to an increase in main clause responses for the Kolkata participants in the within-participants list ( $\beta = 2.61$ ,  $SE = 1.33$ ,  $z = 1.96$ ,  $p = 0.0504$ ). This suggests that participants from Kolkata may have a main clause preference when exposed to both word orders, although the effect of interest persists even in this population.

For the main verb first word order, participants responded with the location denoted by the main verb in 72% of the trials in the within-participants list, and 81% of the trials in the between-participants list. For the embedded verb first word order, participants responded with the location denoted by the main verb in 28% of the within-participants trials, and 19% of the between-participants trials. Thus, we replicated Omaki and colleagues' cross-language findings in the between participant group, and showed a robust bias to resolve the filler-gap dependency with the first verb across word orders in the within participant group as well.

## Discussion

In Experiment 1, we showed that Bangla speakers preferentially resolved a filler-gap dependency with the first position linearly available, regardless of whether this position was in the same clause as the filler or in a more deeply embedded clause. This suggests that the locality constraints determining preferred gap

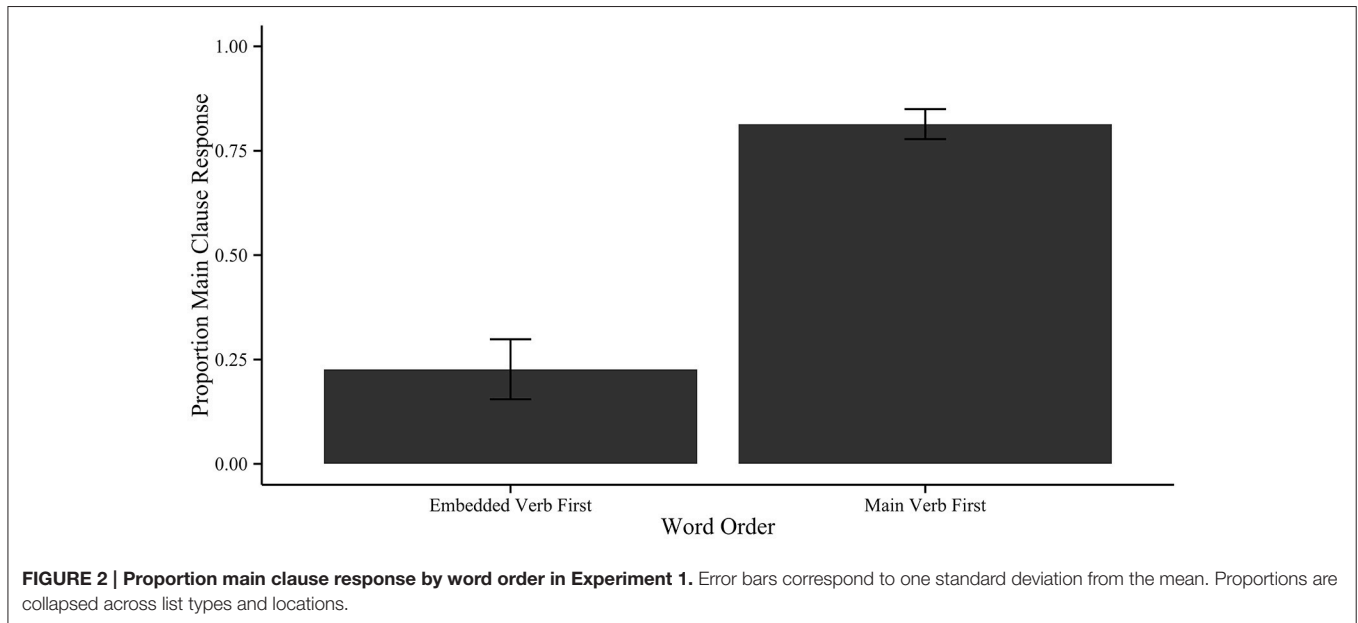
TABLE 2 | Proportion of removed responses in Experiment 1.

	Dhaka		Kolkata	
	Between participants (%)	Within participants (%)	Between participants (%)	Within participants (%)
Main verb first	25	29	31	21
Embedded verb first	15	33	8	23

TABLE 3 | Results of best-fit logistic regression model for Experiment 1.

Fixed effects	Estimate	SE	z	p
(Intercept)	-2.03	0.97	-2.08	0.04*
Word order	5.16	0.95	5.41	<0.001*
City	-1.12	0.92	-1.22	0.22
List type	-0.96	0.91	-1.05	0.29
City * list type	2.62	1.34	1.96	0.05

*P*-values lower than 0,05 are marked with an asterisk.



sites are primarily sensitive to linear distance, as previously shown in a between-language comparison by Omaki et al. (2013). Importantly, this contrasts with observations about island constraints, which appear to be defined in terms of hierarchical structure.

This within-language demonstration of sensitivity to linear order is also important because it helps keep constant all other grammatical properties between the word order comparisons. The results found by Omaki and colleagues may be due to some other grammatical distinction between English and Japanese apart from word order. For instance, obligatory long-distance *wh*-dependencies as observed in English have different properties than the optional *wh*-dependencies observed in Japanese (“scrambling,” Saito, 1985; Mahajan, 1990), which might indirectly bias the filler-gap dependency resolution preferences in these languages. These concerns are less likely to impact the results of Experiment 1, particularly because the effect is robust in the within participant questionnaires. We cannot exclude the possibility that there are subtle formal differences between the pre-verbal and post-verbal filler-gap dependencies. But even if there are such differences, extant accounts of filler-gap dependency processing do not predict that such fine-grained differences should have a large effect on locality biases. We therefore take our findings to lend support to the notion of a general linear locality bias in filler-gap dependency processing.

One potential concern is that the sentences in the embedded-verb first condition may have been parsed as unambiguous. Since the question word *kothae* “where” in (18b) appeared adjacent to the embedded subject it may have been parsed as having a surface position inside the embedded clause. That is, the filler may have been entirely contained in the embedded clause, requiring an embedded clause interpretation. If so, then the embedded clause responses clearly would have been required. However, we consider this unlikely, since these conditions elicited 23% main verb responses, plus additional (excluded) responses in which

participants mentioned both possible answers. So, we think that it is unlikely that these sentences were surface unambiguous for our participants.

An advantage of the Question after Story task in Experiment 1 is that it directly probed participants’ preferred resolution sites instead of measuring whether they detect an unexpected parse, as in the filled-gap effect. However, the Question after Story task does not reveal the time course of dependency formation. We cannot infer from these data that there is early commitment to the linearly first gap site. For this reason, in Experiment 2, we used a filled-gap paradigm in a self-paced reading task to probe for detection of an unsubstantiated gap expectation across word orders.

## EXPERIMENT 2

### Rationale

In Experiment 1, we showed that comprehenders preferred to resolve filler-gap dependencies with the first verb linearly available. The goal of Experiment 2 was to test whether this follows from an early and confident commitment to this gap location. We used the filled-gap paradigm in a self-paced reading task (Crain and Fodor, 1985; Stowe, 1986), which tests whether participants can immediately detect that a previously constructed gap is unavailable. If commitment to the first gap site is made early and confidently enough across word orders, then we expected a filled-gap effect when filler-gap dependency resolution with the first verb was blocked, regardless of whether this occurred in the same clause as the filler or in an embedded clause.

### Participants

Participants were 32 adult native speakers of Bangla from the University of Dhaka student community. Due to a technical error, 3 participants’ responses were not recorded, and thus we report

on 29 participants. They were compensated 500 BDT for their time. The task took approximately 20–30 min to complete.

### Materials

We crossed the factors word order (main verb first or embedded verb first) and extraction type (argument or adjunct extraction). In all target items there was a long-distance *wh*-filler gap dependency. The critical conditions contained an argument *wh*-filler (*ka-ke* “who-ACC”) marked in the accusative case. This argument *wh*-filler was blocked from resolving with the linearly first gap position by a case-matching noun phrase occupying the first canonical object position. This was the filled-gap region and the critical region. The adjunct extraction conditions (*kəkhon* “when,” *kothae* “where”) were the control conditions, since the accusative-marked noun phrase did not block resolution of an adjunct *wh*-dependency in that clause. **Table 4** gives a sample set of items, with the critical filled-gap region underscored and the regions delimited by pipes. Regions were predominantly one word each, except for certain compound verb constructions which contained two words but were treated as one region in the analysis. There were 15 regions in each word order condition, with the filled-gap region being the 7th region in the main-verb first word order, and the 8th region in the embedded-verb first word order.

In the main verb first conditions, the argument *wh*-filler was blocked from resolving as the indirect object of the verb *boleche* “said/told.” The *wh*-filler must then resolve as the direct object of the later, embedded verb. Conversely, in the embedded verb first conditions, the argument *wh*-filler was blocked from resolving as the direct object of the embedded verb. It must therefore resolve as the indirect object of the main verb *boleche*. In the adjunct extraction conditions, an extra pronoun *take* “him/her-ACC” was introduced as the object of the embedded verb in main verb first conditions, and the verb *boleche* “said” in the

embedded verb first conditions. This was necessary to ensure that all verbs had all argument roles discharged. In all conditions, the fronted *wh*-phrase appeared on the left edge of its containing clause to maximize the distance between the *wh*-phrase and the filled-gap region. This prevented the filler from being analyzed as left-adjoined to the embedded clause in the embedded clause first condition. The adjunct *wh*-phrases were counterbalanced between *kəkhon* “when” and *kothae* “where.” The subject of the main clause containing the *wh*-filler always denoted a referent of high status, and the pronoun in the most deeply embedded clause and its verb were morphologically marked with politeness agreement (*tini*). This was done to minimize the complexity induced by any retrieval operations needed in each pronoun and verb region, by maximizing the distinguishability of the referents introduced in the sentence. Additionally, this prevented a potential misanalysis, since a demonstrative is sometimes spelled homographically with the informal third person pronoun (*še*). All target conditions were embedded in an additional clause (*rašad jiggəša koreche...* “Rashad asked...”). This was to ensure that participants could not predict the word order of the target items on the basis of the first few words. There were 32 sets of target items and 48 complexity-matched fillers. The sentences were presented in a Latin Square design, with order randomized for each participant.

### Methods

Sentences were presented on a PC laptop using the Ibox software (<http://www.spellout.net/ibexfarm>) in a self-paced, word-by-word, moving window paradigm (Just et al., 1982). Ibox is intended for web-based tasks, but the task was run offline by one of the authors. Each trial began with a screen presenting a sentence in which the words were masked by dashes, with spaces intact. Each time the participant pressed the spacebar, a word was revealed and the previous word was again hidden behind a dash. A yes/no comprehension question appeared all at once after the participant completed each sentence. The participant was instructed to use the “f” key to respond “yes,” and the “j” key to respond “no,” with on-screen reminders of this key-response pairing. On-screen feedback informed the participant whether the response was correct. Participants were instructed to read carefully at a natural but quick pace, and to answer the questions carefully. The order of presentation of responses was randomized for each participant. All instructions and feedback were given in Bangla.

### Results

Analyses were conducted on comprehension task accuracy and reading times. Trials that received incorrect responses in the comprehension task were removed from analysis. Four participants whose mean accuracy fell below 70% were removed from analysis. The mean accuracy on the comprehension questions was 80.6% after removing these 4 participants.

Using the lme4 package in R (Bates et al., 2015), we analyzed the reading times for the filled-gap region and the subsequent regions using linear mixed effects models for each word order. We included log-transformed reading times as the predicted variable, and extraction type (argument vs. adjunct)

**TABLE 4 | Sample materials from Experiment 2.**

	<i>rašad   jiggəša   koreche</i>   { <i>ka-ke</i>   <i>kothae</i> }   ...	{Argument Adjunct}
	Rashad ask did { <b>who</b>  where} ...	
<b>Main Verb First:</b>	... <u><i>ḍakṭar-ṭa</i></u>   <i>khubi</i>   <i>əbakbhabe</i>   <u><i>rugi-ke</i></u>	
	... doctor-cl very surprisedly patient-acc	
	<i>bolechen</i>   <i>je</i>   <i>tini</i>   <i>purano</i>	
	told that he.pol old	
	<i>haspatal-e</i>   ( <i>take</i> )   <i>cikitša</i>   <i>korechen</i>	
	hospital-loc (him) treatment did	
	“... the doctor told the <u>patient</u> that he treated (him) in the old hospital”	
<b>Embedded Verb First:</b>	... <i>ḍakṭar-ṭa</i>   <i>tini</i>   <i>purano</i>   <i>haspatal-e</i>	
	... doctor-cl he.pol old hospital-loc	
	<u><i>rugi-ke</i></u>   <i>cikitša</i>   <i>korechen</i>	
	patient-acc treatment did	
	<i>bole</i>   <i>khubi</i>   <i>əbakbhabe</i>   ( <i>take</i> )   <i>bolechen</i>	
	that very surprisedley (him) told	
	“... the doctor told (him) that he treated <u>the patient</u> in the old hospital”	

Critical filled-gap region is underlined, and regions are demarcated by pipes.

as the predictor factor. We also included random intercepts for participants and items. For the main verb first conditions, we found no effect of extraction type in the filled-gap region [7th region, *rugi-ke* “patient-ACC,”  $\beta = 0.04$ ,  $SE = 0.04$ ,  $t_{(259)} = 1.0$ ,  $p = 0.32$ ]. However, in the region immediately following the filled-gap region, there was a main effect of extraction type, due to longer reading times in the argument extraction condition [8th region, *boleche* “said,”  $\beta = 0.08$ ,  $SE = 0.04$ ,  $t_{(270)} = 2.1$ ,  $p = 0.04$ ]. This indicates a filled-gap effect for the main verb first word order, suggesting that readers made an early commitment to a gap for the *wh*-filler in this position. Additionally, we found a significant effect of extraction type in the embedded clause, due to longer reading times in the argument extraction condition [12th region, *haspatal-e* “hospital-in,”  $\beta = 66.41$ ,  $SE = 28.87$ ,  $t_{(268)} = 7.3$ ,  $p = 0.02$ ]. This may reflect a secondary filled-gap effect, since it occurs two regions after the embedded clause subject. Comprehenders may have attempted to resolve the filler-gap dependency with the embedded clause subject position. However, this would imply that Bangla speakers do not use case information to determine resolution sites, contrary to our assumptions. Thus, we do not have a good explanation for why reading times should increase at this region. For the embedded verb first conditions, there was no effect of extraction type at the filled-gap region [8th region, *rugi-ke* “patient-ACC,”  $\beta = -0.03$ ,  $SE = 0.04$ ,  $t_{(285)} = -0.69$ ,  $p = 0.49$ ] or in the following two regions [9th region, *cikitsa* “treatment,”  $\beta = 0.05$ ,  $SE = 0.04$ ,  $t_{(302)} = 1.3$ ,  $p = 0.20$ ; 10th region, *korechen* “did,”  $\beta = 0.003$ ,  $SE = 0.004$ ,  $t_{(300)} = 0.08$ ,  $p = 0.94$ ]. Thus, we found no filled-gap effect in the embedded verb first word order. The mean reading times by region are given in **Figures 3, 4**.

## Discussion

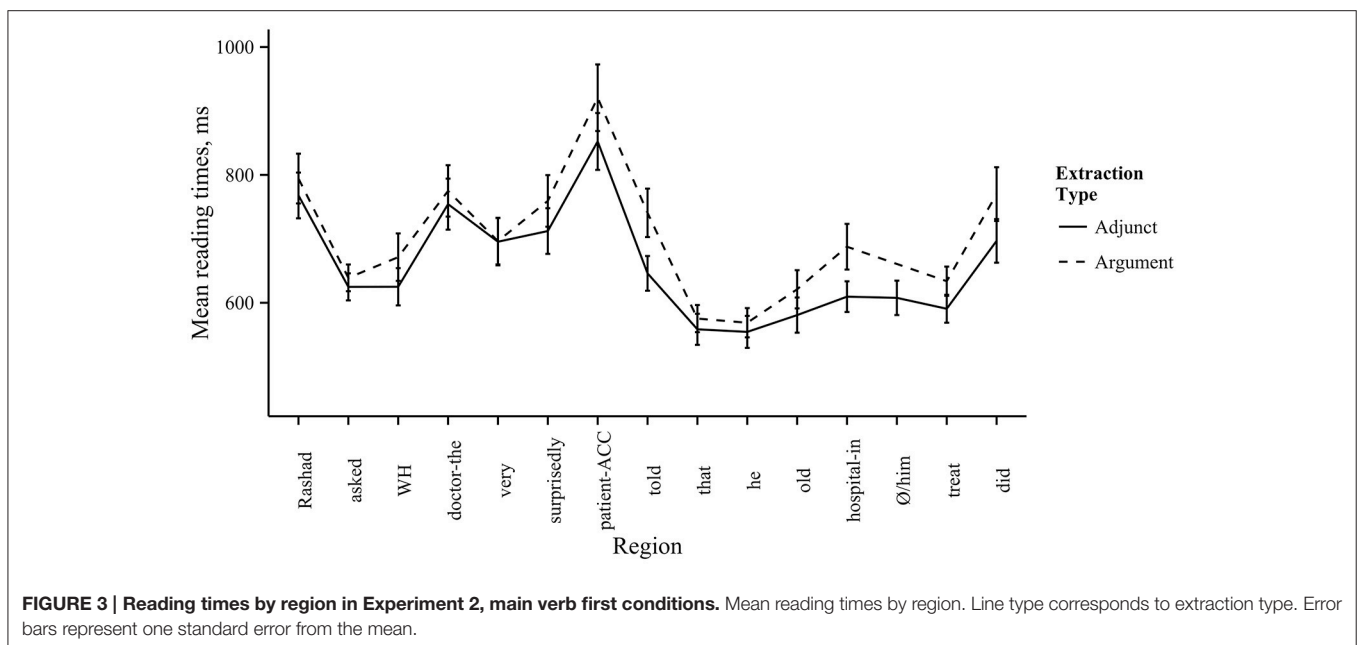
Experiment 2 was designed to probe for sensitivity to blocking of a preferred gap location across word orders using an on-line

measure. If the preference for linearly local gaps found in Experiment 1 reflects an early and confident commitment in both word orders, then we predicted sensitivity to disruption when this resolution was unavailable in the filled-gap paradigm. However, we only found sensitivity to disruption in the main verb first word orders, i.e., when structural locality and linear locality converged. In other words, we did not find evidence of early commitment to this position in Experiment 2. This suggests that the class of gaps that are preferred is not identical to the class of gaps that are committed to early enough to elicit a filled-gap effect.

This difference in measures may be due to a selective sensitivity to structural locality. For instance, the bias to resolve with the first gap linearly available may only manifest as an early commitment when this position is also structurally local. If the biases for preferred gap sites are sensitive to structure in this way, then this undermines one argument for the separation of island constraints from biases on preferred gap sites, i.e., the argument that they should be separated because they refer to different properties of the representation.

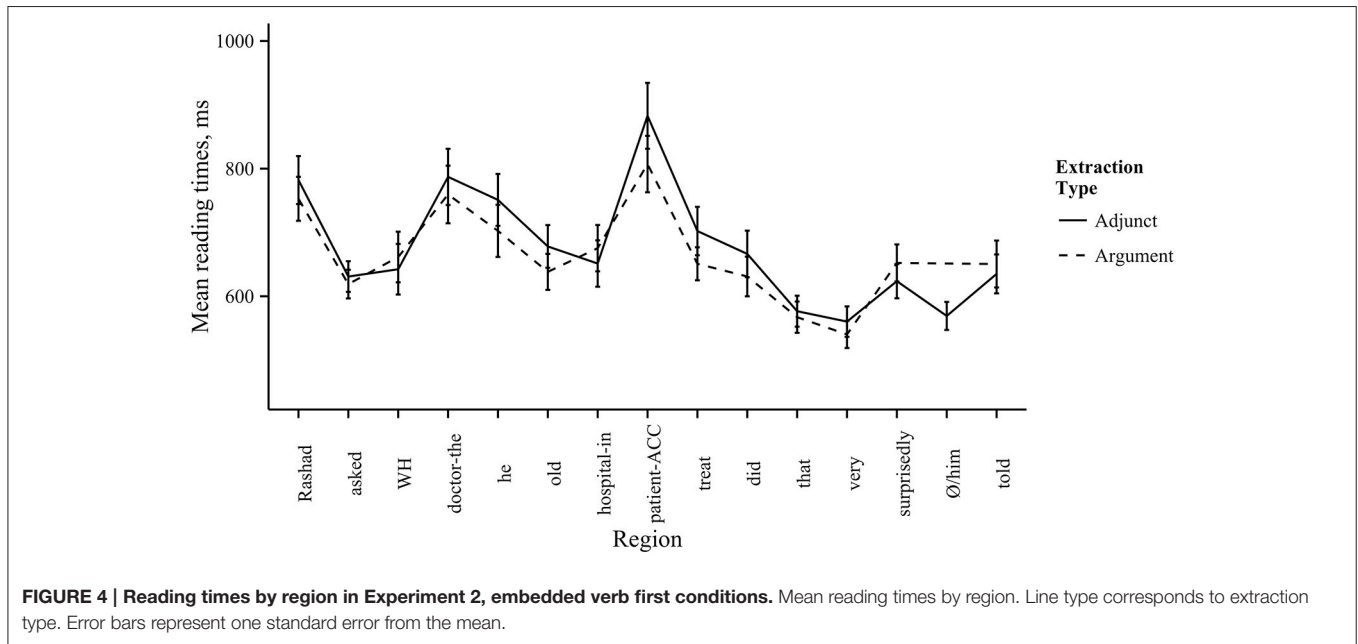
However, the differences between the results in Experiments 1 and 2 may reflect differences between the tasks. Experiment 1 probed directly for resolution preferences in an off-line task. The sentences were globally ambiguous, and somewhat simpler than the three-clause sentences in Experiment 2. Conversely, Experiment 2 was an on-line reading task in which participants read sentences word-by-word.

Participants also seemed to have some difficulty with this task, since their accuracy on the comprehension questions are somewhat lower than average. Additionally, in pilot versions of Experiment 2, participants read at very different paces. Furthermore, participants reported different levels of familiarity with computers, which may have exacerbated the unnaturalness of the task. Some participants struggled with the instructions, e.g., some participants held the space bar down, failing to release it



**FIGURE 3 | Reading times by region in Experiment 2, main verb first conditions.** Mean reading times by region. Line type corresponds to extraction type. Error bars represent one standard error from the mean.





between words. These additional complications in Experiment 2 may have masked an early commitment to resolve a filler-gap dependency with the first gap linearly available, even in embedded verb first word orders. In other words, participants may have made an early commitment to resolve with a linearly local gap site in embedded verb first word orders, but this was selectively masked in Experiment 2.

In Experiment 3, we address these cross-experimental concerns by using the filled-gap paradigm in an offline acceptability judgment task. Experiment 3 was an off-line acceptability judgment task that used two-clause sentences. This was intended to make it as similar to Experiment 1 as possible. The target sentences all contained a filled-gap in a linearly local position, as in Experiment 2. Thus, Experiment 3 relied on an indirect measure of disruption like Experiment 2. Additionally, Experiment 3 was an untimed pen-and-paper task, like Experiment 1, removing the on-line aspect of the self-paced reading paradigm in Experiment 2. If we find evidence of sensitivity to disruption only in main verb first word orders in Experiment 3, i.e. when structural and linear locality converge, then we can conclude that structural locality affects the processes involved in making an early commitment to a gap site. Conversely, if there is a filled-gap effect in Experiment 3 across word orders, then we can infer that the failure to find a filled-gap effect in Experiment 2 was due to the design of that experiment.

## EXPERIMENT 3

### Rationale

In Experiment 3, we again investigated whether Bangla speakers preferred gap sites that are linearly local or structurally local. In Experiment 1, we found evidence for an off-line bias for

linearly local gap sites. In Experiment 2, we found sensitivity to disruption with a filled-gap in a linearly local position, but only with main verb first word order, i.e., when the first gap was also structurally local. In Experiment 3, we investigated whether this mismatch between the results in Experiments 1 and 2 was due to the on-line/off-line contrast between the studies, or the ambiguity resolution/filled-gap paradigm contrast.

Experiment 3 was an off-line acceptability judgment task. In this task, participants read sentences in an untimed way, as in Experiment 1. However, like in Experiment 2, we used a filled-gap paradigm. Although the filled-gap paradigm is typically used in on-line measures, it can also be used to detect filled-gap effects in off-line measures (Sprouse, 2008). This is because the reanalysis associated with detecting a filled-gap effect also lowers ratings in acceptability judgment tasks. Thus, we can compare the ratings for sentences in which the preferred gap is unavailable with controls. If we find a decrease in acceptability, then we take this to be a filled-gap effect. If we find a filled-gap effect across both word orders in Experiment 3, then we can infer that the lack of an embedded clause bias in embedded verb first word orders is due to the design of Experiment 2. Conversely, if we find evidence for a filled-gap effect in the main verb first word order only, then this implies that the difference between Experiments 1 and 2 may be due to the different nature of ambiguity resolution tasks (Experiment 1) and the mechanisms involved in detecting filled-gap effects (Experiments 2–3).

### Participants

Participants were adult native speakers of Bangla drawn from the University of Dhaka and Calcutta University student populations. There were 32 participants from each group. Participants in Dhaka were compensated 500 BDT for their time, and participants in Kolkata were compensated 200 INR. The

experiment lasted approximately 10–20 min, and was conducted after either Experiment 2 or another unrelated experiment.

## Materials

The materials in Experiment 3 were constructed in a similar way to the materials from Experiment 2. We crossed three factors—word order (main verb first or embedded verb first), extraction type (argument extraction or adjunct extraction), and filled-gap position (linearly local or linearly distant). This third factor was added to test for any filled-gap effect with the main clause verb in embedded verb first word orders, i.e., to probe for a filled-gap effect in a position that was linearly distant but structurally local. We constructed 8 lists with an equal number of items per condition, and an equal number of items across lists. There were 24 sets of target items, and 36 complexity-matched fillers, 18 of which were ungrammatical. Each participant saw 3 sentences from each condition and all the fillers in a randomized order.

There were a few differences between the target items in Experiments 2 and 3 that are worth noting. First, the target items in Experiment 3 contained two clauses, unlike the three clause sentences in Experiment 2. This is more similar to the materials in Experiment 1. Additionally, the *wh*-phrase appeared in the pre-verbal position like in Experiment 1, not the left-edge position as in Experiment 2. This was done because the pre-verbal position is perhaps the more canonical position for *wh*-fillers (Simpson and Bhattacharya, 2003). Since the *wh*-filler is in its canonical position adjacent to the embedded clause in embedded verb first word orders, it is possible that comprehenders will treat this as unambiguous, as we suggested in “Section Discussion”. This should bias the results to have an embedded resolution with embedded verb first conditions. Lastly, in the adjunct conditions we did not include the additional object pronoun. In Experiment 2, we included this extra pronoun to ensure that the verb with which the argument *wh*-filler was interpreted had an overt argument in the adjunct extraction conditions. However, this may have been unnecessary, since Bangla permits null arguments. Thus, we did not include this extra pronoun, to maximize

similarity between the argument and adjunct extractions. A sample set of materials is given in **Table 5**.

## Methods

Experiment 3 was a pen-and-paper acceptability judgment study. Participants were instructed to read the sentences carefully, and then circle a number ranging from 1 to 7, with lower scores indicating unacceptability. They were given example sentences with values already circled to illustrate how to use the scale.

## Results

We submitted the ratings to a linear mixed effects model, using the *lme4* package in R (Bates et al., 2015). We included random effects for participant and item. We included word order (main verb first or embedded verb first), extraction type (argument extraction or adjunct extraction), and filled-gap position (linearly local or linearly distant) as predictors, together with their interaction terms. We also included location (Dhaka or Kolkata) in the model. We then used the backwards elimination method to simplify the model using the *step()* function in R, eliminating the location factor. The estimates of the model are presented in **Table 6**. The means of the ratings by condition are given in **Figure 5**. We then performed pairwise comparisons for extraction type within the two word orders and two filled-gap positions, using the least-squares means estimates with Tukey adjustment. These are shown in **Table 7**.

There were two main findings in Experiment 3. First, we found a main effect of word order. Ratings were significantly increased in main verb first word order [ $\beta = 0.87$ ,  $SE = 0.20$ ,  $t_{(1451)} = 4.25$ ,  $p < 0.001$ ]. This is consistent with the observation that main verb first word orders are the preferred word order for clausal embedding in Bangla. Secondly, there was a three-way interaction between word order, filled gap position, and extraction type [ $\beta = -1.39$ ,  $SE = 0.40$ ,  $t_{(1451)} = -3.4$ ,  $p < 0.001$ ]. The only significant pairwise comparison was between argument and adjunct extraction in the main verb first, local filled gap conditions [ $\beta = 1.28$ ,  $SE = 0.20$ ,  $t_{(1450)} = 6.38$ ,  $p < 0.0001$ ]. This

**TABLE 5 | Sample materials for Experiment 3.**

Jahid { <i>ka-ke</i>   <i>kothae</i> } ...	{Argument Adjunct}
Jahid {who-acc where}...	
<b>Main V, Local FG</b>	<p>... <i>khubi</i> <i>ᅇbakbhabe tar bondhu-ke boleche je nipa party-te dekheche</i>            ... very surprisedly his friend-acc said that nipa party-at saw            “Who/where did Jahid very surprisedly tell his friend that Nipa saw at the party?”</p>
<b>Main V, Distant FG</b>	<p>... <i>khubi</i> <i>ᅇbakbhabe boleche je nipa party-te tar bondhu-ke dekheche</i>            ... very surprisedly told that nipa party-at his friend-acc saw            “Who/where did Jahid very surprisedly tell that Nipa saw his friend at the party?”</p>
<b>Embedded V, Local FG</b>	<p>... <i>nipa party-te tar bondhu-ke dekheche bole khubi</i> <i>ᅇbakbhabe boleche</i>            ... Nipa party-at his friend-acc saw that very surprisedly told            “Who/where did Jahid tell very surprisedly that Nipa saw his friend at the party?”</p>
<b>Embedded V, Distant FG</b>	<p>... <i>nipa party-te dekheche bole khubi</i> <i>ᅇbakbhabe tar bondhu-ke boleche</i>            ... Nipa party-at saw that very surprisedly his friend-acc told            “Who/where did Jahid tell his friend very surprisedly that Nipa saw at the party?”</p>

reflects the lowered ratings with main verb first word order, local filled gap, and argument extraction conditions. In other words, there was a decrease in ratings when an argument *wh*-filler could not resolve with the main verb in main verb first word orders. This is a replication of the filled-gap effect in Experiment 2 in off-line acceptability judgments. Crucially, this was only observed in the word orders in which linear and structural locality aligned, i.e., we only found a filled gap effect in situations where the first potential gap position was both structurally and linearly local. Additionally, there was again no difference between participants in Dhaka or Kolkata, suggesting that there is no difference between dialects.

### Discussion

The goal of Experiment 3 was to determine whether Bangla speakers are sensitive to disruption of a linearly local filler-gap dependency resolution using an off-line measure. We conducted Experiment 3 to determine whether the lack of a filled-gap effect

in embedded verb first word orders in Experiment 2 was due to the design of that experiment, or whether it reflects a difference between sensitivity to disruption and general locality preferences, as explored in Experiment 1.

The results from Experiment 3 show that Bangla speakers are only sensitive to disruption of a linearly local filler-gap dependency resolution in main verb first word orders. There was no filled-gap effect with linearly distant filled-gaps in either word order, and there was no filled-gap effect with linearly local filled-gaps in embedded verb first word orders. In other words, we again found a contrast between word orders with respect to sensitivity to disruption with the linearly local gap position.

Thus, we conclude that there is a general bias to resolve filler-gap dependencies with the first position linearly available, but that this only translates into an early and confident commitment when this is also a structurally local position. This means that the bias to resolve with the first position linearly available is only one component of detecting a disrupted filler-gap dependency, as measured with paradigms like the filled-gap effect.

**TABLE 6 | Results of best-fit mixed effects model for Experiment 3.**

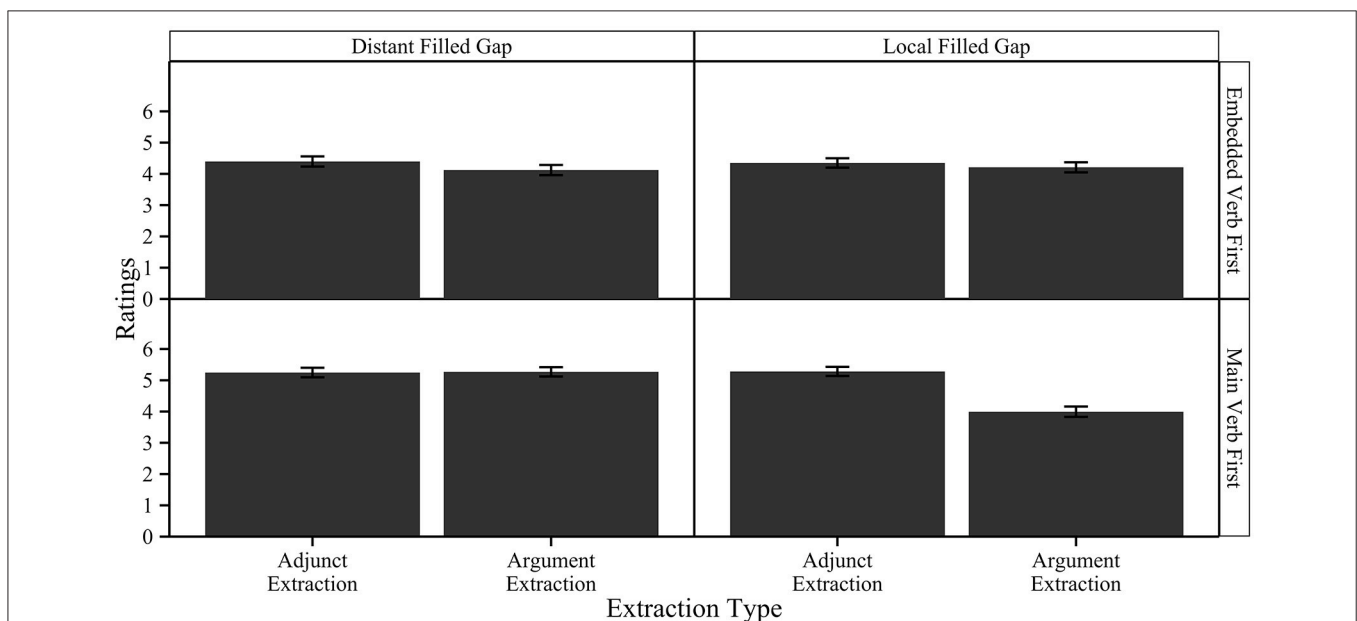
Fixed Effects	Estimate	SE	t	df	p
(Intercept)	4.39	0.18	24.2	281	<0.0001*
Word order	0.87	0.20	4.25	1451	<0.0001*
Filled-gap position	-0.04	0.20	-0.21	1451	0.83
Extraction type	-0.25	0.20	-1.25	1451	0.21
Word order * FGPosition	0.06	0.29	0.21	1452	0.84
Word order * ExtType	0.24	0.29	0.85	1452	0.40
FGPosition * ExtType	0.11	0.28	0.39	51	0.70
WOrder * FGPosition * ExtType	-1.39	0.40	-3.43	1451	0.0006*

*P-values lower than 0.05 are marked with an asterisk.*

**TABLE 7 | Results from pairwise comparisons in Experiment 3.**

Word order, filled gap position	Estimate	SE	t ratio	df	p
Main verb first, local filled gap	1.28	0.20	6.38	1450	<0.0001*
Main verb first, distant filled gap	0.009	0.20	0.043	1436	0.97
Embedded verb first, local filled gap	0.14	0.20	0.70	1428	0.48
Embedded verb first, distant filled gap	0.25	0.20	1.24	1433	0.22

*Comparisons were between argument and adjunct extractions within word order and filled-gap position. Comparisons were made using least squares means with Tukey HSD adjustment. P-values lower than 0.05 are marked with an asterisk.*



**FIGURE 5 | Ratings by condition in Experiment 3.** Error bars represent standard errors of the mean.

## GENERAL DISCUSSION

In this paper, we investigated filler-gap dependency formation in Bangla. Bangla features flexible word order that permits us to manipulate whether the first position linearly available is in an embedded clause or in the main clause. This allows us to manipulate whether linear and structural locality diverge. In main verb first word orders, the first gap position linearly available is also structurally local, whereas in embedded verb first word orders the first gap position linearly available is structurally distant. In Experiment 1, we found a strong bias to resolve an ambiguous filler-gap dependency with the first position linearly available, regardless of its structural depth. However, in Experiments 2 and 3, we only found evidence of sensitivity to disruption when linear and structural locality converge. We interpreted these results as showing that there is a general bias for linearly local gaps, but that this only translates into a strong early commitment to this gap site when the linearly local gap is also structurally local.

If we start with the assumption that preferred gaps are typically detectable in a filled-gap paradigm, then the finding that these measures diverge for embedded verb first word orders in Bangla is surprising. The filled-gap paradigm, an indirect measure of gap formation preferences that we used in Experiments 2 and 3, depends on multiple processing mechanisms. The paradigm requires that participants make an early and confident commitment to a gap site, presumably in accordance with their linear locality preferences. Upon encountering the filled gap, the comprehender must quickly detect that the preferred gap is unavailable, and then instigate a costly reanalysis process. The lack of a filled gap effect in embedded verb first word orders might be attributed to any of these processes failing to deploy quickly.

Bangla speakers may not have shown a filled-gap effect in embedded verb first conditions in Experiments 2 and 3 because this word order is dispreferred. This was reflected in the lowered ratings for this word order in Experiment 3. This may be in part because pre-verbal embedded clauses have specific semantic and syntactic restrictions, unlike post-verbal embedded clauses (e.g., Bayer, 2001). In our estimation, long pre-verbal embedded clauses are also likely less frequent in naturalistic speech, and may carry certain pragmatic or discourse functions that also make them atypical. As a consequence, Bangla speakers may find processing pre-verbal embedded clauses more difficult, and have less facility making fine-grained predictions in pre-verbal embedded clauses for that reason. This contrasts with Japanese, in which pre-verbal embedded clauses are canonical (Tanaka, 2001), and filled-gap effects are found in pre-verbal embedded clauses (Aoshima et al., 2004; Yoshida, 2006).

Another salient difference between Bangla and Japanese is the case system. In Japanese, nominative marking surfaces as a morpheme *-ga*, and is used to quickly detect embedded clauses in real-time processing (Miyamoto, 2002). However, Bangla nominative noun phrases are morphologically unmarked. We speculated in Section Grammatical Properties of Bangla that Bangla speakers should be able to detect an embedded

clause in embedded verb first word orders by observing two animate noun phrases, zero-marked for nominative. This is because case-marking morphemes are obligatory for animate noun phrases in non-subject positions. However, Bangla speakers might not compute this immediately. If Bangla speakers cannot immediately detect the embedded clause nominative subject as such, then construction of the embedded clause might be delayed, potentially even until the embedded verb. The relative timing of the construction of the embedded clause could explain the difference between these languages and the two measures in Bangla. Aoshima et al. (2004) suggested that the embedded clause bias in Japanese follows from a reanalysis triggered by the onset of the embedded clause. They propose that a Japanese comprehender first commits to a main clause gap position for a filler-gap dependency, and then revises to an embedded clause interpretation upon detecting the embedded clause. Crucially, this means that the comprehender has committed to a gap site in the embedded clause before encountering the filled-gap. If Bangla speakers cannot detect the embedded clause until after the filled gap, then there is no commitment to a gap resolution in the embedded object position by the time the comprehender enters that region. Thus, Bangla speakers should not show any evidence of reanalysis in the filled gap region. On this view, we predict no filled-gap effect in this context (as observed in Experiment 2), nor any reduction in judgments associated with such a reanalysis (as observed in Experiment 3). However, by hypothesis, Bangla speakers still have an embedded clause preference, and in off-line tasks they eventually select an analysis where a filler-gap dependency resolves in this position, if the string permits it. Thus, measures that directly probe for preferences in ambiguity resolution are predicted to reveal an embedded clause resolution preference when this position is linearly first, as in Experiment 1. A clear prediction of this account is that a language that has word order flexibility like Bangla but a case marking system like in Japanese should exhibit filled-gap effects in pre-verbal embedded contexts. An explanation that leverages these differences in information flow due to differences in case-marking may be the most promising framework for explaining these apparent differences.

Another possibility is that the contrast between our experiments is due to differences between argument and adjunct *wh*-dependencies. Experiment 1 tested adjunct *wh*-questions, whereas Experiments 2 and 3 tested argument *wh*-questions for the target conditions. This was by design, because adjuncts more easily permit the crucial ambiguity in Experiment 1, and argument *wh*-questions are more amenable to the filled-gap paradigm. However, there is little existing evidence that argument and adjunct *wh*-dependencies are comprehended differently. If resolving filler-gap dependencies is motivated by the need to find a semantic role for the unintegrated filler, then this should be the case regardless of the type of filler (e.g., Pritchett, 1992). Unpublished work has demonstrated filled-gap effects for adjunct *wh*-phrases (Yoshida and Dickey, 2008), and recent work suggests an increase in processing difficulty associated for sentences with adjunct *wh*-phrases compared to sentences with no filler-gap dependency, implying active search

processes are used even for adjunct filler-gap dependencies (Stepanov and Stateva, 2015).

Finally, another possible confound in our results is that Experiment 1 had rich contexts presented before the target item, but Experiments 2 and 3 did not. It is possible that this may have impacted the results in some way. However, we balanced the materials such that both interpretations of the ambiguous *wh*-question were pragmatically plausible. Additionally, findings from the Question After Story task in Omaki et al. (2013) English and Japanese converge with on-line findings in similarly context-less reading tasks (Aoshima et al., 2004). Thus, we tentatively take the results from the Question after Story task to reveal the same biases in filler-gap dependency processing as are observed in reading context-less sentences. However, it remains possible that context plays a greater role in Bangla than it does in prior Japanese studies.

## CONCLUSION

Much work in theoretical linguistics and psycholinguistics demonstrates that there are robust locality constraints on gaps. Both structural locality and linear locality play important roles in selecting gaps in real-time sentence processing. Structural locality is relevant for determining which gap sites are grammatically well-formed, and linear locality is relevant for determining which gaps are preferred when multiple potential gaps are available (Aoshima et al., 2004; Omaki et al., 2013). Locality biases on filler-gap dependencies can reveal themselves in different ways—as a general preference for certain gap sites, or as an early commitment. Typically, these are taken to reflect the same processes of active dependency formation, but different measures show that they dissociate. We investigated the dissociation of linear locality and structural locality by manipulating the flexible word order of Bangla, which allows testing the contribution of structural and linear locality.

In Experiment 1, we showed that Bangla speakers have a preference for linearly local gaps, regardless of structural position. This replicated findings from a previous English and Japanese comparison within the same language (Omaki et al., 2013), and thus supports the generalization that filler-gap dependency locality preferences are primarily sensitive to linear locality. However, in Experiments 2 and 3, we found evidence for filled-gap effects only when the disrupted position

was both the linearly first position in the sentence and structurally closest. We highlighted a few reasons why this difference between word orders might hold. Specifically, we suggested that gaps in pre-verbal embedded clauses may be difficult to maintain, because of the status of pre-verbal embedded clauses in Bangla. Alternatively, we suggested that the informativity of the case system is such that comprehenders may not have a commitment to a gap position in place before the filled-gap region in the embedded verb first word orders. These facts contrast with Japanese, which exhibits a strong bias for gaps in pre-verbal embedded clauses (Aoshima et al., 2004; Yoshida, 2006; Omaki et al., 2013). If the results from Experiments 2 and 3 are amenable to these kinds of explanations, then it may be possible to retain the hypothesis that linear locality determines preferred gap sites in filler-gap dependency processing, whereas structural locality determines acceptable gap sites.

## AUTHOR CONTRIBUTIONS

DC: primary author, designed and conducted experiments; MI: primary author, designed and conducted experiments; SD: secondary author, designed and conducted experiments; SM: secondary author, designed experiments; MD: secondary author, designed experiments; CP: secondary author, designed experiments.

## ACKNOWLEDGMENTS

The authors would like to thank Awalín Sopan, Titir Santra, and Fuad Hasan Sobuj for assistance in preparing the materials and recruiting participants. We previously presented these results at the 26th and 27th CUNY Sentence Processing conferences, and the 3rd and 4th Formal Approaches to South Asian Languages conference. This work was funded by grant NSF-GRFP grant DGE-0750616 awarded to DC, NSF grant NSF BCS-0848554 awarded to CP, and NSF grant DGE-0801465 awarded to UMD.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01235>

## REFERENCES

- Abrusán, M. (2011a). Presuppositional and negative islands: a semantic account. *Nat. Lang. Semantics* 19, 257–321. doi: 10.1007/s11050-010-9064-4
- Abrusán, M. (2011b). *Wh*-islands in degree questions: a semantic approach. *Semant. Pragmat.* 4, 1–44. doi: 10.3765/sp.4.5
- Alexopoulou, T., and Keller, F. (2007). Locality, cyclicity, and resumption: at the interface between the grammar and the human sentence processor. *Language* 83, 110–160. doi: 10.1353/lan.2007.0001
- Altmann, G. T. M., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Ambridge, B., and Goldberg, A. (2008). The island status of clausal complements: evidence in favor of an information structure explanation. *Cogn. Linguist.* 19, 357–389. doi: 10.1515/COGL.2008.014
- Aoshima, S., Phillips, C., and Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *J. Mem. Lang.* 51, 23–54. doi: 10.1016/j.jml.2004.03.001
- Bal, B. K. (1990). *Comp and Complementizers in Oriya and English*. Ph.D. dissertation. Central Institute of English and Foreign Languages.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). *lme4: Linear Mixed-Effects Models Using Eigen and s4*. Available online at: <http://CRAN.R-project.org/package=lme4>.
- Bayer, J. (1996). *Directionality and Logical Form*. Boston, MA: Kluwer.

- Bayer, J. (1999). Final complementizers in hybrid languages. *J. Linguist.* 35, 233–271. doi: 10.1017/S0022226799007665
- Bayer, J. (2001). “Two grammars in one: sentential complements and complementizers in Bengali and other South-Asian languages,” in *The Yearbook of South Asian Linguistics and Languages: Tokyo Symposium on South-Asian languages – Contact, Converge, and Typology*, eds P. Bhaskaro and K. V. Subbarao (New Delhi: Sage Publications), 11–36.
- Boeckx, C. (2008). Islands. *Lang. Linguist. Compass* 2, 151–167. doi: 10.1111/j.1749-818X.2007.00043.x
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., and Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: evidence from *wh*-questions. *J. Mem. Lang.* 34, 774–806. doi: 10.1006/jmla.1995.1034
- Bourday, J. S. (1992). “Parsing complex NPs in French,” in *Island Constraints: Theory, Acquisition, and Processing*, eds H. Goodluck and M. S. Rochemond (Dordrecht: Kluwer), 61–87.
- Chomsky, N. (1977). “On *wh*-movement,” in *Formal Syntax*, eds P. Culicover, T. Wasow, and A. Akmajian (New York, NY: Academic Press), 71–133.
- Chomsky, N. (1981). *Lectures in Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Barriers*. Cambridge: MIT Press.
- Cowart, W. (1996). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.
- Cowart, W. (2003). “Detecting syntactic dialects: the that-trace phenomenon,” in *Paper Presented at the 39th Chicago Linguistic Society*. Chicago, IL.
- Crain, S., and Fodor, J. D. (1985). “How can grammars help parsers?” in *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*, eds D. Dowty, L. Karttunen, and A. Zwicky (Cambridge: Cambridge University Press), 94–128.
- Dasgupta, P. (2007). “The ubiquitous complementizer,” in *Linguistic Theory and South Asian Languages: Essays in Honor of KA Jayaseelan*, eds J. Bayer, T. Bhattacharya, and M. T. Hany Babu (Amsterdam: John Benjamins), 163–173.
- David, A. B. (2015). *A Descriptive Grammar of Bengali*. New York, NY: De Gruyter Mouton.
- Dayal, V. (1996). *Locality in WH Quantification*. Dordrecht: Kluwer.
- Deane, P. D. (1991). Limits to attention: a cognitive theory of island phenomena. *Cogn. Linguist.* 2, 1–63. doi: 10.1515/cogl.1991.2.1.1
- de Villiers, J., Roeper, T., and Vainikka, A. (1990). “The acquisition of long-distance rules,” in *Language Processing and Language Acquisition*, eds L. Frazier and J. de Villiers (Dordrecht: Kluwer), 257–297.
- Dickey, M. W. (1996). Constraints on the sentence processor and the distribution of resumptive pronouns. *Linguist. Lab.* 19, 157–192.
- Erteschik-Shir, N. (1973). *On the Nature of Island Constraints*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Faraway, J. J. (2002). *Practical Regression and Anova in R*. Available online at: <http://www.maths.bath.ac.uk/~jjf23/book/>
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguist. Inq.* 9, 427–473.
- Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Nat. Lang. Linguist. Theory* 5, 519–559. doi: 10.1007/BF00138988
- Frazier, L., and Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Lang. Cogn. Process.* 4, 93–126. doi: 10.1080/01690968908406359
- Frazier, L., and Flores d’Arcais, G. B. (1989). Filler driven parsing: a study of gap filling in Dutch. *J. Mem. Lang.* 28, 331–344.
- Garnsey, S. M., Tanenhaus, M. K., and Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *J. Psycholinguist. Res.* 18, 51–60. doi: 10.1007/BF01069046
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Heestand, D., Xiang, M., and Polinsky, M. (2011). Resumption still does not rescue islands. *Linguist. Inq.* 42, 138–152. doi: 10.1162/LING\_a\_00032
- Hofmeister, P., and Sag, I. (2010). Cognitive constraints and island effects. *Language* 86, 366–415. doi: 10.1353/lan.0.0223
- Huang, C.-T. J. (1982). *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Jackendoff, R. S., and Culicover, P. (1971). A reconsideration of dative movements. *Found. Lang.* 7, 397–412.
- Just, M. A., Carpenter, P. A., and Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *J. Exp. Psychol.* 111, 228–238. doi: 10.1037/0096-3445.111.2.228
- Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 is an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084
- Kluender, R. (1998). “On the distinction between strong and weak islands: a processing perspective,” in *Syntax and Semantics: The Limits of Syntax*, eds P. Culicover and L. McNally (New York, NY: Academic Press), 241–280.
- Kluender, R. (2004). “Are subject islands subject to a processing account?,” in *Proceedings of the 23rd West Coast Conference on Formal Linguistics (WCCFL 23)*, eds V. Chand, A. Kelleher, A. Rodriguez, and B. Schmeiser (Somerville, MA: Cascadia Press), 475–499.
- Kluender, R., and Kutas, M. (1993). Subjacency as a processing phenomena. *Lang. Cogn. Process.* 8, 573–633. doi: 10.1080/01690969308407588
- Kuno, S. (1976). “Subject, theme, and speaker’s empathy – a reexamination of relativization phenomena,” in *Subject and Topic*, ed C. Li (New York, NY: Academic Press), 417–444.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package)*. Available online at: <http://CRAN.R-project.org/package=lmerTest>
- Lasnik, H., and Saito, M. (1992). *Move Alpha: Conditions on Its Application and Output*. Cambridge: MIT Press.
- Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *J. Psycholinguist. Res.* 33, 51–73. doi: 10.1023/B:JOPR.0000010514.50468.30
- Lewis, M. P., Simons, G. F., and Fennig, C. D. (2015). *Ethnologue: Languages of the World, 18th Edn*. Dallas, TX: SIL International.
- Lewis, S., and Phillips, C. (2015). Aligning grammatical theories and language processing models. *J. Psycholinguist. Res.* 10, 206–233. doi: 10.1007/s10936-014-9329-z
- Mahajan, A. (1990). *The A/A-bar Distinction and Movement Theory*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Manetta, E. (2012). *Peripheries in Kashmiri and Hindi/Urdu*. Amsterdam: John Benjamins.
- Manzini, M. R. (1992). *Locality: A Theory and Some of Its Empirical Consequences*. Cambridge: MIT Press.
- McElree, B. (2006). “Accessing recent events,” in *The Psychology of Learning and Motivation, Vol. 46*, ed B. H. Ross (San Diego, CA: Academic Press), 155–200.
- McElree, B., and Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence processing: a time-course analysis. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 432–460. doi: 10.1037/0278-7393.24.2.432
- McKinnon, R., and Osterhout, L. (1996). Event-related potentials and sentence processing: evidence from the status of constraints on movement phenomena. *Lang. Cogn. Process.* 11, 495–523. doi: 10.1080/016909693871032
- Miyamoto, E. T. (2002). Case markers as clause boundary inducers in Japanese. *J. Psycholinguist. Res.* 31, 307–347. doi: 10.1023/A:1019540324040
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., and Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *J. Cogn. Neurosci.* 3, 151–165. doi: 10.1162/jocn.1991.3.2.151
- Nicol, J. L., Fodor, J. D., and Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1229–1238. doi: 10.1037/0278-7393.20.5.1229
- Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *J. Psycholinguist. Res.* 18, 5–19. doi: 10.1007/BF01069043
- Omaki, A., Davidson-White, I., Goro, T., Lidz, J., and Phillips, C. (2013). No fear of commitment: children’s incremental interpretation in English and Japanese *wh*-questions. *Lang. Learn. Dev.* 10, 206–233. doi: 10.1080/15475441.2013.844048
- Phillips, C. (2006). The real-time status of island phenomena. *Language* 82, 795–823. doi: 10.1353/lan.2006.0217
- Phillips, C. (2013). “On the nature of island constraints I: language processing and reductionist accounts,” in *Experimental Syntax and Island Effects*, eds J. Sprouse and N. Hornstein (Cambridge: Cambridge University Press), 64–108.

- Phillips, C., Kazanina, N., and Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cogn. Brain Res.* 82, 795–823. doi: 10.1016/j.cogbrainres.2004.09.012
- Pritchett, B. L. (1992). *Grammatical Competence and Parsing Performance*. Chicago, IL: University of Chicago Press.
- Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris.
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge: MIT Press.
- Rizzi, L. (2013). Locality. *Lingua* 130, 169–186. doi: 10.1016/j.lingua.2012.12.002
- Ross, J. R. (1967). *Constraints on Variables in Syntax*. Ph.D. dissertation, Massachusetts Institute of Technology.
- Rudin, C. (1988). On multiple questions and multiple *wh*-fronting. *Nat. Lang. Linguist. Theory* 6, 445–501. doi: 10.1007/BF00134489
- Saito, M. (1985). *Some Asymmetries in Japanese and Their Theoretical Implications*. Ph.D. thesis, MIT.
- Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition* 71, 109–147. doi: 10.1016/S0010-0277(99)00025-6
- Simpson, A., and Bhattacharya, T. (2000). “*WH-CP pied piping in Bangla*,” in *Proceedings of the 30th Annual Meeting of the North-East Linguistic Society*, eds M. Hirotani, A. Coetzee, N. Hall, and J.-Y. Kim (Amherst, MA: GLSA), 583–596.
- Simpson, A., and Bhattacharya, T. (2003). Obligatory *wh*-movement in a *wh-in-situ* language. *Linguist. Inq.* 34, 127–142. doi: 10.1162/002438903763255940
- Sobin, N. (1987). The variable status of comp-trace phenomena. *Nat. Lang. Linguist. Theory* 5, 33–60. doi: 10.1007/BF00161867
- Sprouse, J. (2008). The differential sensitivity of acceptability to processing effects. *Linguist. Inq.* 39, 686–694. doi: 10.1162/ling.2008.39.4.686
- Sprouse, J., Wagers, M., and Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language* 88, 82–123. doi: 10.1353/lan.2012.0004
- Srikumar, K. (2007). “Clausal pied-piping and subjacency,” in *Linguistic Theory and South Asian Languages: Essays in Honor of KA Jayaseelan*, eds J. Bayer, T. Bhattacharya, and M. T. Hany Babu (Amsterdam: John Benjamins), 53–69.
- Stepanov, A., and Stateva, P. (2015). Cross-linguistic evidence for memory storage costs in filler-gap dependencies with *wh*-adjuncts. *Front. Psychol.* 4:1301. doi: 10.3389/fpsyg.2015.01301
- Stowe, L. (1986). Evidence for on-line gap location. *Lang. Cogn. Process.* 1, 227–342. doi: 10.1080/01690968608407062
- Sussman, R. S., and Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: evidence from eye-movements during spoken *wh*-questions. *Lang. Cogn. Process.* 18, 143–163. doi: 10.1080/01690960143000498
- Szabolcsi, A., and den Dikken, M. (1999). Islands. *GLOT Int.* 4, 3–8.
- Szabolcsi, A., and Zwarts, F. (1993). Weak islands and algebraic semantics of scope-taking. *Nat. Lang. Semantics* 1, 235–284. doi: 10.1007/BF00263545
- Tanaka, H. (2001). Right-Dislocation as scrambling. *J. Linguist.* 37, 551–579.
- Tanenhaus, M. K., Stowe, L. A., and Carlson, C. N. (1985). “The interaction of lexical expectation and pragmatics in parsing filler-gap constructions,” in *Seventh Annual Cognitive Science Society Conference* (Hillsdale: Erlbaum), 361–365.
- Thompson, H. R. (2010). *Bengali: A Comprehensive Grammar*. London: Routledge.
- Traxler, M. J., and Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: an eye-tracking study. *J. Mem. Lang.* 35, 457–475. doi: 10.1006/jmla.1996.0025
- Truswell, R. (2007). Extraction from adjuncts and the structure of events. *Lingua* 117, 1355–1377. doi: 10.1016/j.lingua.2006.06.003
- Uriagereka, J. (1992). “The syntax of movement in Basque,” in *Syntactic Theory and Basque Syntax*, eds J. Ortiz de Urbina and J. A. Lakarra (Bilbao: Euskal Herriko Unibersitatea), 417–445.
- Wagers, M. W., and Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *J. Linguist.* 45, 395–433. doi: 10.1017/S002226709005726
- Wagers, M. W., and Phillips, C. (2014). Going the distance: memory and control processes in active dependency construction. *Q. J. Exp. Psychol.* 67, 1274–1304. doi: 10.1080/17470218.2013.858363
- Wanner, E., and Maratsos, M. (1978). “An ATN approach in comprehension,” in *Linguistic Theory and Psychological Reality*, eds M. Halle, J. Bresnan, and G. Miller (Cambridge: MIT Press), 119–161.
- Weinberg, A. S. (1992). Parameters in the theory of sentence processing: minimal commitment theory goes east. *J. Psycholinguist. Res.* 3, 339–364.
- Yoshida, M. (2006). *Constraints and Mechanisms in Long-Distance Dependency Formation*. Ph.D. dissertation, University of Maryland.
- Yoshida, M., and Dickey, M. W. (2008). “Active dependency formation for adjunct *wh*-phrases,” in *Poster Presented at CUNY 2008* (New York, NY).
- Yoshida, M., Kazanina, N., Pablos, L., and Strut, P. (2014). On the origin of islands. *Lang. Cogn. Neurosci.* 29, 761–770. doi: 10.1080/01690965.2013.788196

**Conflict of Interest Statement:** 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 © 2016 Chacón, Imtiaz, Dasgupta, Murshed, Dan and Phillips. 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) or licensor 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.



# Experience-Based Probabilities Modulate Expectations in a Gender-Coded Artificial Language

Anton Öttl\* and Dawn M. Behne

Speech Lab, Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University–Stillwater,  
USA

### Reviewed by:

Itziar Laka,  
University of the Basque Country,  
Spain  
Rachel Helen Messer,  
Oklahoma State University–Stillwater,  
USA

Viktória Havas,  
Norwegian University of Science and  
Technology, Norway

### \*Correspondence:

Anton Öttl  
anton.ottl@ntnu.no

### Specialty section:

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

**Received:** 31 October 2015

**Accepted:** 05 August 2016

**Published:** 23 August 2016

### Citation:

Öttl A and Behne D (2016)  
Experience-Based Probabilities  
Modulate Expectations in a  
Gender-Coded Artificial Language.  
*Front. Psychol.* 7:1250.  
doi: 10.3389/fpsyg.2016.01250

The current study combines artificial language learning with visual world eyetracking to investigate acquisition of representations associating spoken words and visual referents using morphologically complex pseudowords. Pseudowords were constructed to consistently encode referential gender by means of suffixation for a set of imaginary figures that could be either male or female. During training, the frequency of exposure to pseudowords and their imaginary figure referents were manipulated such that a given word and its referent would be more likely to occur in either the masculine form or the feminine form, or both forms would be equally likely. Results show that these experience-based probabilities affect the formation of new representations to the extent that participants were faster at recognizing a referent whose gender was consistent with the induced expectation than a referent whose gender was inconsistent with this expectation. Disambiguating gender information available from the suffix did not mask the induced expectations. Eyetracking data provide additional evidence that such expectations surface during online lexical processing. Taken together, these findings indicate that experience-based information is accessible during the earliest stages of processing, and are consistent with the view that language comprehension depends on the activation of perceptual memory traces.

**Keywords:** artificial language, frequencies of exposure, mental representation, visual world eyetracking, gender representations, experience-based probabilities

## 1. INTRODUCTION

The present study investigates the acquisition and subsequent processing of new associations between a spoken, morphologically complex word and a visual referent. The aim of the study is to assess whether experience-based probabilities are actively used to predict an upcoming referent, or whether such information is overshadowed by other disambiguating information. If experience-based probabilities are actively used, what are the temporal dynamics of processing? Taken together, these questions have implications for understanding the activation of probabilistic information during online language processing, and more generally how referential meaning is (re)constructed during comprehension.

### 1.1. Words, Referents, and Conceptual Representations

Upon hearing a word like “swan” in isolation, we are more likely to picture a white swan than a black one. Even if the word does not explicitly encode information about the color attribute, this undeniably constitutes a perceptually salient feature of the bird the word typically refers



to. Compared to a black swan, a white swan is consistent with the swans we most frequently encounter, and therefore represents a more likely referent, and also a more typical instantiation of the underlying concept (e.g., Rosch, 1978).

Over the last decades, the view that perceptually based information plays a central role in the cognitive processing of conceptual representations (Barsalou, 1999; Zwaan, 2004; Richter et al., 2009; Richter and Zwaan, 2010) has gained acceptance. For example, Zwaan et al. (2002) presented participants with sentences like “*The ranger saw the eagle in the sky*” before showing them a picture of an eagle, and asking them to indicate whether this had been mentioned in the preceding sentence. If the depiction of the eagle matched the implied shape (eagle with spread wings), response times were shorter than when it did not (eagle with folded wings). This indicates that mental representations constructed during language comprehension are highly specified in terms of perceptually based information, i.e., the mental representation of an event incorporates information that is not explicitly encoded in the input, but which is available from experience. Numerous studies report similar findings for other perceptual dimensions such as shape, orientation and color information (see Zwaan and Pecher, 2012 for a discussion and recent replication experiments), and provide additional evidence that representations are constructed incrementally (Sato et al., 2013), unconsciously (Pecher et al., 2009a; Vukovic and Williams, 2014) and are not driven by task dependent strategies (Pecher et al., 2009b). In contrast to more traditional views of concepts as amodal and abstract entities, perception-based accounts typically envision conceptual representation as a dynamic simulation process. Barsalou (1999) proposes that perceptual memory traces stored in long-term memory take on a symbolic function as they come to analogically represent referents in communication, i.e., the neural substrate that is activated during the perception of an object overlaps with its conceptual representation.

Similarly, when we process words referring to human beings, we activate expectations as to whether these refer to females or males. For example, upon hearing a word like “*nurse*,” we might be more likely to picture a woman than a man. Being a non-arbitrary classification based on psychologically salient features, gender categorization offers a fertile ground to further investigate how experience-based information is reflected in mental representations associated with words.

## 1.2. Gender Representations in Natural Language Processing

An expectation that “*pilot*” refers to a male person may be based on prior experience with more pilots being male than female. However, associative links between words and their previously encountered referents are not likely to be simplistic. First, associations may be indirect in the sense that gender expectations are derived from other aspects of meaning (e.g., that the role of a pilot is “masculine” regardless of the actual proportion of male pilots), or also from the coactivation of related concepts (*aeroplanes*, *engineering*, or *Ray-Bans*) that also bear the potential of activating gender associated information. Second, a given concept likely integrates information from

various sensory modalities (e.g., the expected gender distribution among pilots may be more dependent on auditory than visual information, assuming that pilots are more often heard over in-flight audio systems than they are actually seen in person). Third, expectations may be partially or fully derived from secondary sources (be they news reports or fiction). Finally, role nouns may be used non-referentially, generically (i.e., leaving referential gender open) or without fine grained information about their intended reference becoming revealed. Studies of gender expectations in natural language processing therefore typically investigate stereotype-based expectations as available from ratings of role nouns (cross-linguistic norming data on the conceptual gender of role nouns are available in Misersky et al., 2013).

A substantial body of research indicates that processing difficulties occur when stereotypical gender information associated with a given word is not consistent with its reference in a given context. The experimental paradigms under which these effects have been observed typically address the activation of conceptual information indirectly by looking at the implications it has for subsequent processing, either in priming tasks or in sentence contexts. For example, Oakhill et al. (2005) asked participants to judge whether a role noun (e.g., “*footballer*”) could refer to the same person as a previously presented kinship term (e.g., “*sister*”), and found a response facilitation for word pairs where the stereotypical gender of the target word was congruent with the primed kinship term (see also Banaji and Hardin, 1996). Similar effects have also been found in sentence and story contexts (Garnham et al., 2002; Reynolds et al., 2006). Recent research further shows that these effects are not restricted to the use of stereotyped role nouns, but are also found when the stereotypical information is conveyed by descriptions of gender typical activities (Reali et al., 2015), such as the description “*repairs and produces furniture*.” This latter point is of particular relevance, it suggests that stereotype effects may depend on the activation of complex knowledge structures, and not merely on the direct associations between nouns referring to humans and the gender of likely referents.

The visual world paradigm, in which gazes toward different images in a visual display are recorded while the participant is presented with auditory information, has been used successfully to examine online language processing in the absence of additional judgment tasks (for a review of the paradigm see Huettig et al., 2011), and has been applied to investigate whether gender stereotypical information is activated automatically even when it is not relevant to the experimental task (Pyykkönen et al., 2010). In their study, Pyykkönen et al. presented Finnish participants with auditorily recorded short stories containing gender-stereotypical role nouns while looking at a display showing a male and a female character along with two additional story elements. Even if the gender of the characters was not needed to establish discourse coherence, the character whose gender was consistent with the noun stereotype was fixated significantly more than the character whose gender was inconsistent with the stereotype. While this provides strong evidence that gender stereotypical information is automatically triggered by the noun, the analysis collapsed fixations over a

relatively large time-window that contained additional linguistic information. Therefore, the results do not specify whether stereotype information was accessed immediately during lexical processing of the role noun, or if it was only inferred after a more abstract and definitional meaning associated with the noun had been retrieved, and/or if the effect depended on the additional linguistic context.

Taken together, the studies mentioned in this section (Banaji and Hardin, 1996; Garnham et al., 2002; Oakhill et al., 2005; Reynolds et al., 2006; Pyykkönen et al., 2010) demonstrate that gender stereotypical information is automatically activated during online language processing. The present study contributes to this line of research by investigating the formation and processing of expectations at the lexical level in more detail. More specifically, we examine whether stereotype-based expectations observed in natural language processing may reflect relatively shallow semantic processing (i.e., frequency weighted associations between words and referents as opposed to requiring more complex semantic processing to occur). In parallel, we explore how these expectations unfold over time, both on the timescale of acquisition and on the timescale of online lexical processing. If the effects surface during lexical processing, this would imply that gender information is intrinsically linked to more definitional conceptual attributes, as opposed to being inferred from a more abstract representation later in processing. To our knowledge, these aspects remain unexplored in the literature, and would have important implications for understanding the cognitive-linguistic representation of gender, but also for understanding the relationship between perceptual experience and cognitive-semantic representations more generally.

### 1.3. Current Experiment

In order to overcome some of the complexities associated with naturally occurring conceptual representations discussed in the preceding sections (e.g., that concepts are linked to related concepts and are likely based on information obtained from numerous sources) as well as the variability found in natural language (e.g., differences in wordlength would make it difficult to investigate aspects of online processing), the present study adopts an artificial language. Acquiring artificial language materials in an experimental setting differs from natural language acquisition in numerous ways (e.g., in terms of communicative settings and goals, the duration of learning sessions and the complexity of materials acquired), and necessarily comes at some cost to ecological validity (see e.g., Hulstijn, 1997 for a discussion). Nevertheless, as demonstrated by previous research, such as word segmentation and category-based abstraction (see e.g., Gómez and Gerken, 2000 for a review), artificial languages can be used to tap into the cognitive resources that a learner brings to the task. Artificial language experiments to some extent represent an idealized learning situation, and findings from such experiments must be seen as supplementary to findings from experiments using natural language materials.

By training the acquisition of pseudowords and their reference to a set of imaginary figures, a simplified referential system is established, in which words are linked to visual referents,

and both words and referents vary on a controlled and limited number of dimensions. Pseudowords were constructed to encode gender by means of suffixation, such that a wordstem would denote the overall features of an imaginary figure whereas the suffix would denote its gender, mimicking structures found in natural languages (see Section 2.3 for more details). Encoding gender information in this manner allows for referential ambiguity during the processing of the stem while also providing disambiguating gender information from the suffix. Based on previous research with artificial lexicons (Magnuson et al., 2003), we expect the artificial language to be successfully acquired within a relatively short experimental session, and also that the acquired words will be processed similarly to natural words to the extent that the gender coding system is comparable to corresponding systems found in natural language(s).

## 2. METHODS

### 2.1. Design

The present experiment consists of three parts: (a) a pre-test in which participants are familiarized with the stimuli to be acquired, (b) a training phase in which participants learn the new word-referent associations, and (c) a post-test in which the processing of the newly acquired representations is evaluated. As outlined in detail in Section 2.4, experience-based expectations were induced during the training phase. Throughout the experiment, the participant's task is to identify which of four alternative images a given word refers to. Although participants will start by guessing in the pretest, the training phase should lead to the establishment of new associative links in memory, which will be explored in the post-test. For each response, accuracy and response time was collected, and eyetracking was used to measure which image was looked at in the course of each trial.

### 2.2. Participants

Twenty-three native speakers of Norwegian (12 male, mean age = 24.3,  $SD = 2.7$ ) were recruited at the Norwegian University of Science and Technology (NTNU) in Trondheim. All participants reported normal hearing and normal or corrected-to-normal vision and were compensated for participation with a gift certificate. Participants were naïve to the critical aspect of the experiment (i.e., the induction of probability-based expectations). They gave their written informed consent by using a form approved by the Data Protection Official for Research for Norwegian universities (NSD).

### 2.3. Materials

Pseudowords were constructed to consist of two elements: a word stem and a suffix. For each pseudoword, the stem would refer to an imaginary figure whereas the suffix would indicate its gender. Out of the vast possibilities to encode gender, this system allows control both of word lengths and the exact timepoint at which gender information becomes available, both of which are critical for the exploration of the online processing of spoken language stimuli. Structurally, these pseudowords would be comparable to English wordforms like “*policeman*”/“*policewoman*,” although the pseudowords (a) control for word-length, (b) avoid generic

usage (e.g., the word “policeman” is sometimes used with a female referent), and (c) avoid the use of free morphemes (e.g., “police”) with independent semantic associations. The current experiment was conducted with native speakers of Norwegian. Like English, Norwegian is a relatively gender neutral language to the extent that gender information is typically not encoded at the lexical level (exceptions do exist: “*servitrise*” and “*politimann*” are structurally and semantically equivalent to English “*waitress*” and “*policeman*,” but are increasingly replaced by gender neutral alternatives). In contrast to English, Norwegian is a grammatical gender language [e.g., grammatically masculine “*bilen*” (English: *the car*), vs. grammatically neuter “*huset*” (English: *the house*)]. Importantly, for words referring to human beings, grammatical gender is not linked to referential gender, and most role nouns are grammatically masculine. Thus, while the gender-coding system underlying the artificial language is not likely to be immediately transparent to the participants, its overall structure should not be exotic. For an overview over gender distinctions across languages see e.g., Corbett (1991).

### 2.3.1. Auditory Stimuli

The artificial lexicon was designed to encode gender through suffixation: twelve pseudoword stems were paired with two different pseudosuffixes (“-tef” and “-tok”) to comprise a lexicon consisting of 24 items (see **Figure 1** for a schematic representation, or Appendix A for a full overview). All pseudowords were controlled in terms of consonant-vowel patterns, phonotactic probabilities and lexical neighborhood densities. Audio recordings of the pseudowords were spoken by a young adult female native speaker of Urban East Norwegian (Kristoffersen, 2000) and recorded with a Røde NT1-A microphone at a sampling rate of 44.1 kHz in Praat version 5.3 (Boersma, 2001). The speaker was seated in a sound attenuated booth, and presented with the words in randomized order on a computer display. In total, approximately ten tokens of each pseudoword were recorded, and these were independently evaluated by two raters, taking into consideration background noise, breathing or any other disruptive feature that seemed relevant. Experimental stimuli were selected from the highest ranked tokens, with particular attention to their similarity in intonation and speech rate.

Previous research has demonstrated that fine acoustic-phonetic detail can be actively used to predict upcoming information during online processing at the lexical level (Salverda et al., 2003). For the current study, this implies that participants could theoretically exploit acoustic-phonetic cues from the word stem to predict whether it would end in “-tef” or “-tok,” obscuring the true timepoint at which information about referential gender was available. Therefore, the 24 original recorded tokens were cross-spliced, i.e., audiofiles (e.g., “*bontok*” and “*bontef*”) were cut at the syllable boundary to obtain separate audiofiles for stems and suffixes (e.g., “*bon*,”<sup>a</sup> “*tok*,”<sup>a</sup> “*bon*,”<sup>b</sup> “*tef*”<sup>b</sup>) which were then recombined to produce additional tokens (e.g., “*bon*<sup>a</sup>*tef*,”<sup>b</sup> “*bon*<sup>b</sup>*tok*”<sup>a</sup>) in Praat (Boersma, 2001). The final stimulus pool thus contained two tokens of each word (i.e., 48 in total) that were used interchangeably. On average, pseudowords had a duration of 865 ms ( $SD = 71$ ). For each word, the

uniqueness point (in terms of specifying which of the 24 figure it refers to) was defined as the onset of the vowel in the second syllable: for words ending in “-tef” the average uniqueness point was 440 ms after onset ( $SD = 46$ ); for words ending in “-tok” 440 ms after onset ( $SD = 47$ ). Thus, the earliest point at which it would be theoretically possible to unambiguously identify the referent of a given word would be approximately 400 ms after its onset. This latter estimate is deliberately somewhat conservative to acknowledge the variance in stem-durations, and thereby ensure that effects triggered during the processing of the suffix are not erroneously detected in the time-window that primarily reflects the processing of the stem.

### 2.3.2. Visual Stimuli

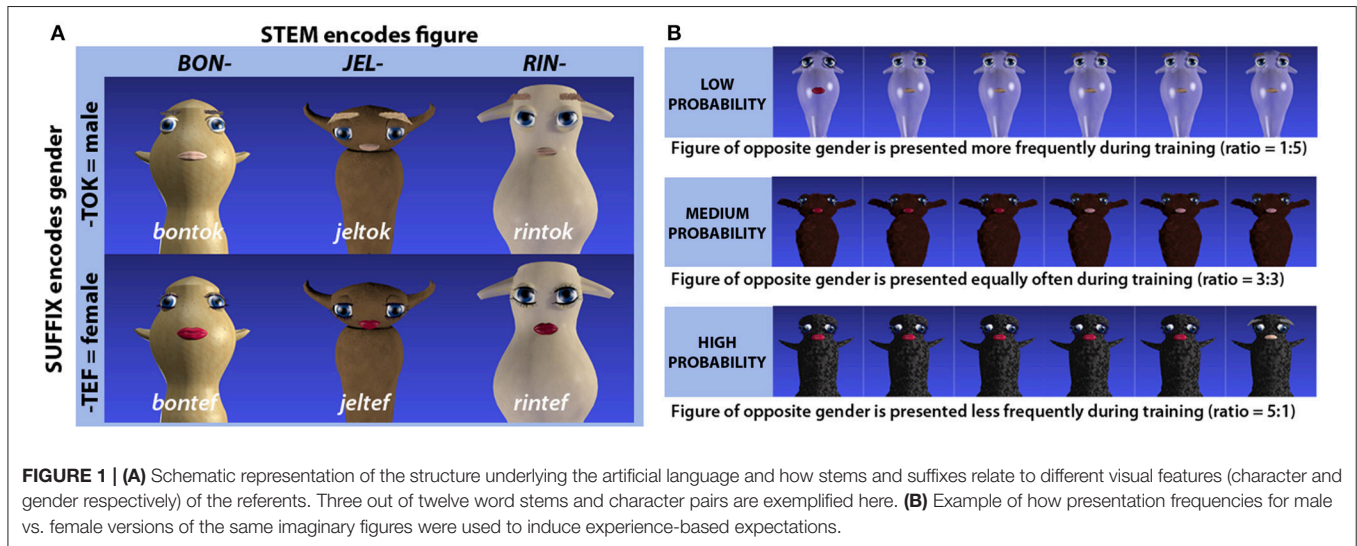
The visual stimuli were designed to provide referents for the artificial language outlined above, and depicted imaginary figures that could be either male or female (the full set is included in Appendix B, along with additional information about the structuring). To keep the image set as symmetric as possible, twelve base figures without any cues to gender were created in a first step. The global shapes of these figures were based on novel object stimuli (courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, <http://www.tarrlab.org>), but were modified extensively for the current experiment. To make the figures easily distinguishable from each other, the following global features were used: overall shape, color and surface texture (e.g., shiny, furry, matte). In a second step, female and male versions of these figures were obtained by adding salient and consistent gender cues: female figures were given red lips and long eye-lashes whereas male figures were given lighter but short eye-lashes, slightly smaller pink lips and bushy eyebrows. Importantly, all gender cues can be considered local features of the facial region, in contrast to the global features distinguishing the different base figures from each other. Thus, shared features between two members of a figure pair should be more salient than the features that distinguish them according to gender, as would arguably be the case for similar representations in naturalistic settings. All images were created using Blender 3D modeling software, version 2.60 (Blender Foundation, 2012), and are available from the corresponding author upon request.

### 2.3.3. Sound—Image Associations

Each of the twelve word stems was consistently linked to one of the twelve base figures, while the two suffixes were consistently linked to the gender identity of a given figure (see **Figure 1**). The links between word stems and base figures were randomly assigned for different participants, and for one half of the participants, the suffix “-tok” was assigned to male and “-tef” to female figures, while for the other half, the gender assignment was reversed.

## 2.4. Procedure

To induce experience-based probabilities, frequencies of exposure to male vs. female realizations of the same auditorily presented words and associated visual imaginary figures were manipulated during training. For one third of the word-referent



pairs, the female version appeared as a target five times while the male version appeared as a target only once per training block. For another third, the male version appeared five times and the female version once per block. For the remaining word-referent pairs, female and male versions appeared as targets equally often (3 times each per block). Thus, according to the presentation frequencies, each item would have a low, medium or high probability of being associated with a given gender, the bias-strengths being 18.75, 50, and 81.25%, respectively (see **Figure 1**). In the following, these groups of items will be referred to as low, medium or high probability targets.

Testing took place in a sound attenuated booth in the Speech Lab at the Department of Psychology at NTNU. Participants were seated approximately 70 cm from a computer display. Eprime 2.0.8.90 was used to run the experiment and a SmartEye 5.8 remote system was used for the collection of gaze data (at a sampling frequency of 60 Hz), with SmartEye extension for Eprime (Version 1.0.1.49) to handle the communication between the two. Auditory stimuli were presented over AKG MKII K271 headphones and responses were collected using a computer mouse connected to the stimulus PC. The experiment was controlled from outside the booth.

Testing consisted of a pre-test (24 trials), five training blocks (72 trials each) and a post-test (144 trials). The experiment duration was approximately 1 h, with an additional 15–30 min for among other things calibration. Participants were informed about the overall structure of the experiment (outlined in detail below) prior to participation, and were also reminded at the transition between each part of the experiment.

#### 2.4.1. Pre-test

Participants were informed that they would be familiarized with the words and images that they would acquire in the course of the experiment, that they would see four characters on the screen, listen to a nonsense word, and then have to guess which of the images the word belonged to by clicking with the mouse on

one of the images. Crucially, they were also instructed about the gender coding: they were told which gender each suffix denoted, and were presented with two examples to familiarize them with visual gender cues. Each trial began with a fixation cross in the center of the display. When this had been fixated for 500 ms, four images (two male and two female figures) appeared on the display. After the four images had been displayed 500 ms, a pseudoword corresponding to one of these was presented over the headphones, while the images remained on the display. Once a response had been made, a gray frame appeared around the selected image to indicate that the response had been registered. After 500 ms, all images were removed from the display. If no image was selected within 4500 ms, the experiment would automatically move on to the next trial. Each of the 24 stimuli appeared once as a target and three times as a distractor. Image displays were randomized, but never featured the male and the female version of the same base figure at the same time. Nor would the same word stem appear as a target in two consecutive trials. At the end of the pre-test participants received feedback as to how many percent of their answers were correct, and were informed that they would now proceed to the training.

#### 2.4.2. Training Blocks

Participants were told that the task would be the same as in the pre-test, but that they would receive feedback after each response whether they had selected the correct image or not. As soon as a participant had selected one of the images, this would receive a green frame if the response was correct, or a red frame if the response was incorrect. After 500 ms (or 4500 ms after word onset, if no image had been selected), the incorrect images were removed from the display, while the correct image remained and the pseudoword was repeated over the headphones. Randomization procedures were identical to the pre-test. If a cross-spliced token was used for the task, the original token was used for the feedback and vice versa. In contrast to the pre-test where each figure appeared as a target

once, presentation frequencies to male vs. female realizations of the same base figures (and correspondingly the associated stem-suffix combination) were manipulated to create three different probability groups, as outlined above. Each training block was followed by feedback on percent correct responses, and a 30-s break. After the final training block, participants were informed that they would now proceed to the post-test.

### 2.4.3. Post-test

The trial structure and randomization procedures were identical to the pre-test. However, the visual displays in the post-test differed from both the pre-test and the training blocks in that three different trial types (within participants) were used to investigate different aspects of processing (Figure 2). One trial type was identical to the pre-test, and always contained four unrelated images. This was defined as a **no competitor trial**, since any target image could unambiguously be identified by the word stem and the global features alone (e.g., a target image associated with the word “gontef” would be accompanied by three distractor images associated with unrelated words “sjestok,” “kestef,” and “lentok,” rendering both the suffix and the visual gender cues redundant). A second trial type contained one image associated with the same base figure as the target word, but of the opposite gender, and was defined as a **target competitor trial**. For example, based on Figure 2, if the target word was “gontef,” the image associated with “gontok” would be among the distractors, and the target and competitor would be distinguishable only by the suffix and the local gender features. Finally the third trial type, defined as a **distractor competitor trial**, featured two distractors differing only in gender features. The inclusion of the latter trial type was deemed necessary to prevent participants from adopting a response strategy according to which the presence of both the male and the female version of a figure would narrow down the response options prior to onset of the word. In the post-test, all words/figures appeared twice as a target in each of the three trial types, regardless of whether they belonged to the low, medium or high probability group.

## 3. RESULTS AND DISCUSSION

We expect participants to acquire the 24 words within the training blocks provided. Since the gender coding was made explicit to the participants, we expect participants to consistently select a figure of the correct gender (as denoted by the suffix), already during the pre-test. Regarding response times for successfully acquired word-referent pairs, we expect participants to be quicker at recognizing high-probability items than low-probability items, provided that probabilistic information is tracked during acquisition, and also readily available during recognition. If experience-based probability effects cannot be detected in the response times, this would suggest that participants indeed activate more abstracted (gender-less) representations. Such a scenario would be likely if participants primarily relied on the information that was encoded linguistically. Finally, regarding the gaze data, we expect the stimulus materials to be sufficiently well acquired for language-driven patterns such as lexical competition effects to be observed. Crucially, we also expect the gaze data to provide information on how quickly probabilistic information becomes available during online processing.

Prior to the analyses, participants' use of gender information was assessed in the data collected during the pretest. One participant selected the figure of the correct gender in only 58% percent of the pretest trials, which suggests that this participant had not paid attention to the instructions at the beginning of the experiment. This participant was excluded from further analyses. The remaining 22 participants (11 male) all scored above 85% correct on the gender coding in the pretest (mean = 95%, 95% CI[93, 97]). As for overall acquisition of the word-image associations (i.e., including correct identification of the stem), performance was at 47% correct (95% CI[42, 51]) in the pretest. As participants were aware of the gender coding, this corresponds to chance performance. In the final training block performance was at 95% correct (95% CI[92, 99]), which indicates that the image-word correspondences were successfully acquired in the provided training blocks.

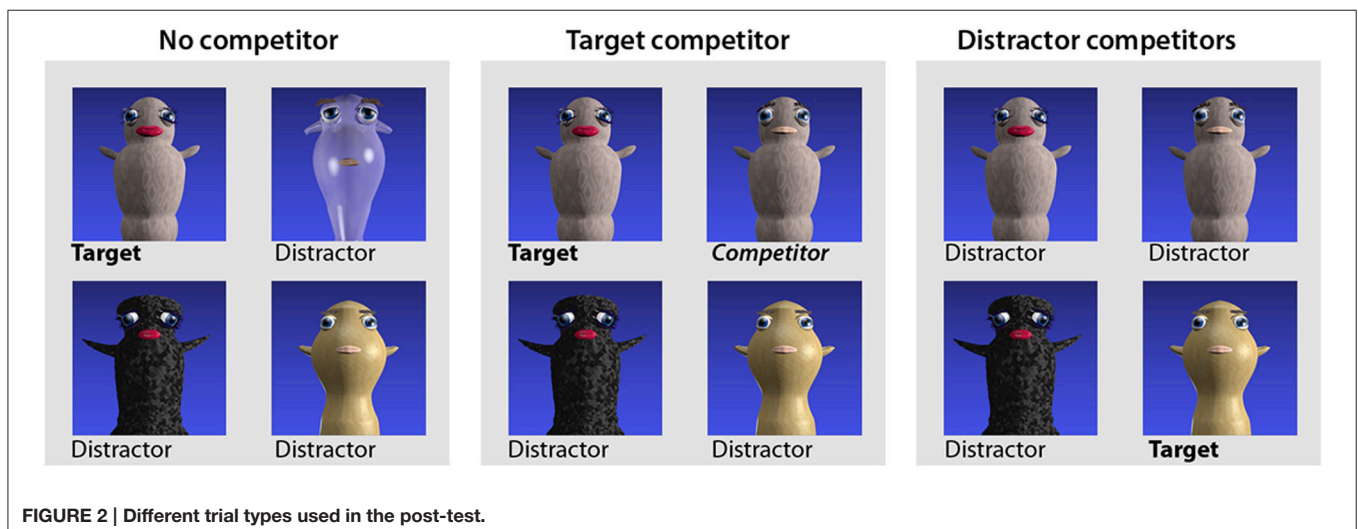


FIGURE 2 | Different trial types used in the post-test.

### 3.1. Statistical Procedures

The analyses of data from the post-test presented below are based on linear mixed effects models in R, version 3.0.2 (R Core Team, 2013), using the *lmer* and *glmer* functions (depending on the dependent variable being continuous or binomial) from the *lme4* package (Bates et al., 2014). Model comparisons were performed using log likelihood tests, using a forward-testing approach: fixed effects are included one at a time, and their contribution to improving model fit is evaluated by comparing the respective model to one that is identical except for not containing the fixed effect in question. Model comparisons to arrive at the best fitting model are included in the Supplemental Materials Data sheet 2. In line with current recommendations (Barr et al., 2013), maximal random effects structure as justified by the design was used, i.e., in addition to random by-subject intercepts, random by-subject slopes were included for the fixed effects being tested. The models did not include random intercepts for items due to the randomization procedures (outlined in Section 2.3): each participant acquired different word-referent associations randomly assigned to the three probability groups, and items would therefore not constitute a natural grouping factor comparable to natural words. The contribution of random slopes to the model fit was also assessed using log likelihood tests. The inclusion of the relevant random slopes was warranted for all models (i.e., for trial type and probability in the response time analysis, and for time and probability in the analyses of gaze data). To obtain *p*-values for the best fitting models, *lme4* was used in conjunction with the *lmerTest* package (Kuznetsova et al., 2014).

When trial type is included as a fixed effect in a model, the intercept represents no competitor trials, and this estimate can be directly compared to the adjustments required for target competitor trials and distractor competitor trials. Correspondingly, when probability is included as a fixed effect, the intercept represents low probability items and direct comparisons to medium and high probability items are available from the model estimates. When both effects are included in the same model, the intercept represents the estimate for low probability items in no competitor trials. To facilitate interpretation in cases where an interaction is not included in the model, the relevant estimates for the given factor are reported in isolation, using percentages instead of log likelihoods and milliseconds instead of log transformed milliseconds. To test differences between the factor levels that can not be read directly from the model (i.e., between medium and high probability items or between target competitor and distractor competitor trials), the factor levels were reordered prior to recalculating the same model, in line with recommendations outlined in Singer and Willett (2003).

### 3.2. Response Times

As the auditory word durations were controlled and all word stems could be distinguished based on the phonetic onsets alone (i.e., each stem began with a different consonant), response times were measured from the onset of the word. Only correct responses that were longer than 300 ms were analyzed (95% of the data). Responses made earlier than this are more likely to

be erroneous button presses than to reflect actual recognition (e.g., Baayen, 2008). The response times did not follow a normal distribution, hence a log transformation was performed prior to conducting the analyses (as suggested in Baayen, 2008, a.o.).

For target competitor trials, response times are expected to be longer compared to no competitor trials, since participants need to await disambiguating auditory information from the suffix in order to identify the target. For distractor competitor trials, the presence of two similar images could be distracting in its own right (leading to slower responses), or on the contrary, it could be used as an opportunity to adopt a strategy according to which two response alternatives can be eliminated as soon as the stem has been identified (leading to quicker response times). If experience-based probabilities affect the processing of the newly acquired representations, this is expected to surface as longer response times for low probability items and/or shorter response times for high probability items, reflecting relative ease of processing.

The best fitting model includes fixed effects for trial type and probability, but not their interaction term. The estimates obtained from this model are summarized in **Table 1**, and the aggregated data are presented in **Figure 3**. When no competitor is present, the response time is estimated at 1799 ms (95% CI[1675, 1932]). Relative to this, response times were significantly longer in target competitor trials (1874 ms, 95% CI[1813, 1938]), and significantly shorter in distractor competitor trials (1743 ms, 95% CI[1699, 1789]). Compared to the overall response time for low probability items (1799 ms, 95% CI[1675, 1932]), response times were shorter both for medium probability items (1772 ms, 95% CI[1696, 1852]) and high probability items (1703 ms, 95% CI[1662, 1745]), but only the latter contrast was significant.

These results indicate that experience-based probabilities affected recognition times. Increased processing times when identifying a target whose gender was inconsistent with the induced expectation suggests that gender information is inherent to the newly formed representations. Had participants relied on more general representations (i.e., based on the linguistically encoded information), this difference in response times would be hard to explain. Although the results from the response times do not pinpoint the timepoint at which probabilistic information affected processing, this issue will be explored in the following section, and discussed in more detail in the general discussion.

**TABLE 1 | Model estimates for response time data (logarithmic scale).**

	Estimate	SE	df	t-value	Pr(>t)
(Intercept)	7.495	0.037	21	205.79	<0.001
<b>TRIAL TYPE</b>					
Distractor comp.	-0.031	0.013	70.3	-2.39	<0.05
Target comp.	0.041	0.017	24.6	2.42	<0.05
<b>PROBABILITY</b>					
Medium	-0.015	0.022	21.1	-0.65	0.521
High	-0.055	0.013	435.2	-4.37	<0.001

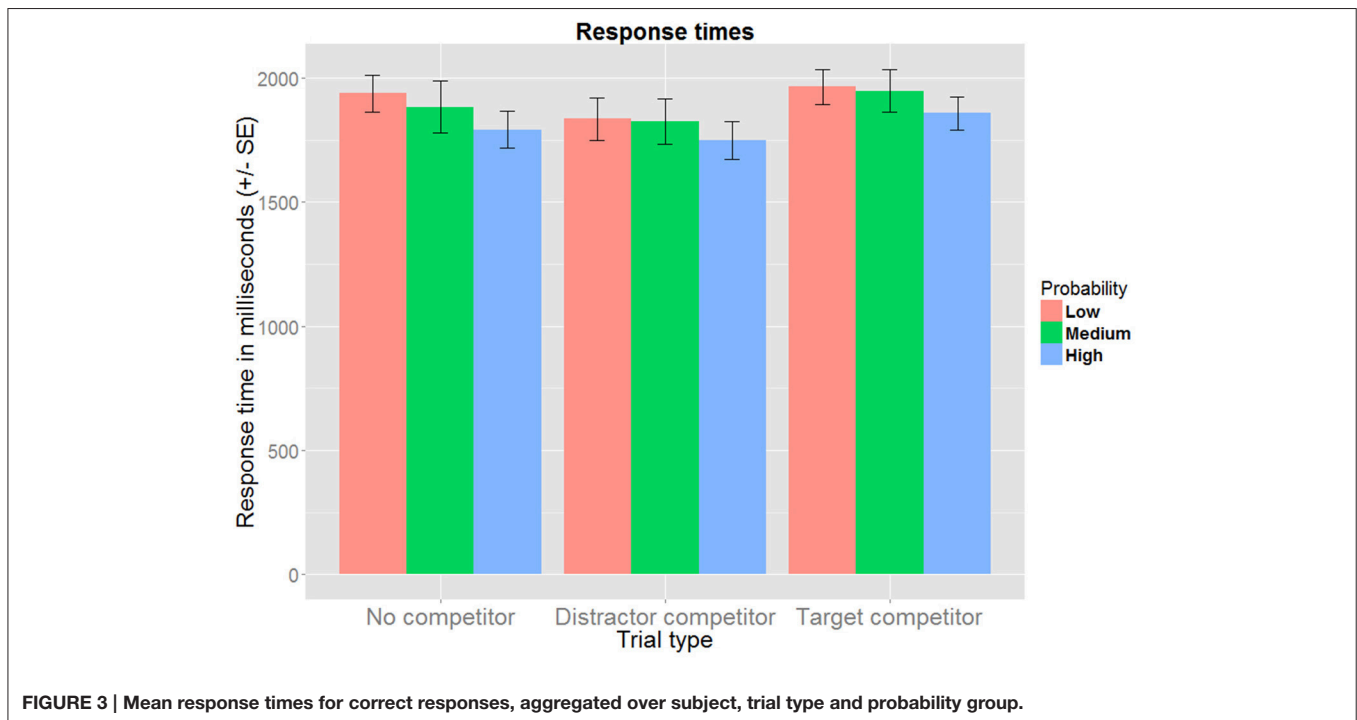


FIGURE 3 | Mean response times for correct responses, aggregated over subject, trial type and probability group.

### 3.3. Gaze Data

Gaze data were collected during the entire post-test, and provide a continuous record of which of the four images in the display was fixated during each trial, from the point when the images appeared on the display until a response had been made. Each obtained gaze coordinate was classified as falling within one of four regions of interest (corresponding to the four image positions on the display), or as falling outside these regions. These data were analyzed to investigate whether the effect of experience-based probabilities found for the response times is observed during online processing, and if so, at which point in time it would emerge.

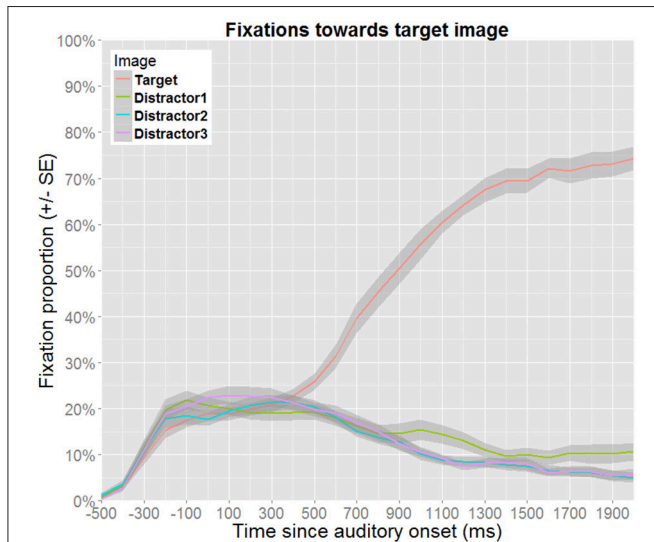
#### 3.3.1. Overall Fixation Patterns in Relation to Auditory Information

The planning and execution of an eye movement is estimated to require approximately 200 ms (e.g., Matin et al., 1993), which implies that the gaze data reflect a delayed response to the auditory information. Hence, in order to compensate for such delays, the time windows of analysis are shifted to begin and end 200 ms after the acoustic onsets and offsets of the events of interest, as is common for this paradigm (e.g., Huettig and Altmann, 2005). **Figure 4** presents the overall fixation patterns for no competitor trials, and **Figure 5** for target competitor trials. In both cases, the gaze patterns show that the auditory information is used incrementally to identify the target.

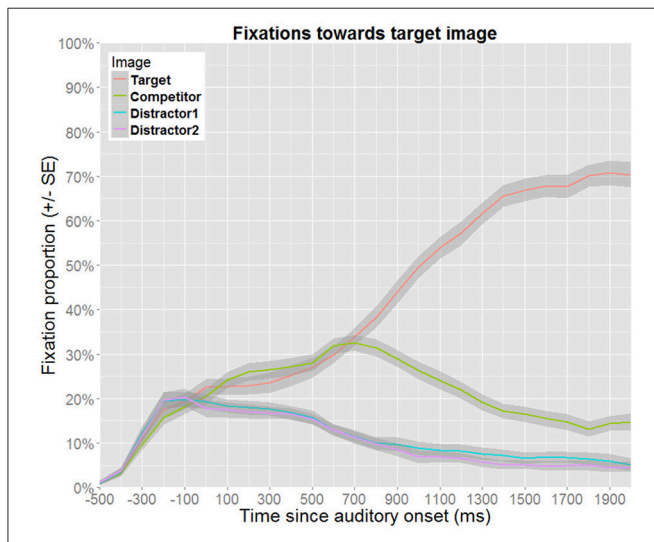
Two epochs of the timeline are of particular interest. First, probabilistic information may be available from the word stem, and such an effect can be expected to be detectable in the

time-window ranging from 200 to 600 ms after onset of the target word, i.e., corresponding to the time between the onset of the word stem and the onset of the suffix. Second, probabilistic information may also be triggered while processing the suffix (e.g., for a given word stem, one suffix may be expected while the other is unexpected), and the second time-window of interest is therefore defined as the range from 600 to 1000 ms after the onset of the word.

The two time-windows of interest were analyzed separately, since this allows time to be modeled as a linear predictor, which facilitates model specification, estimation and interpretation. The analysis of no competitor trials is followed by a separate analysis of target competitor trials. Incorrect responses are excluded from all analyses (for trials without a competitor: 4.5% of the data, for trials featuring a target competitor: 4.6% of the data). To retain maximal temporal resolution, unaggregated data were used for the analyses (visualization in **Figures 6, 7** use aggregated data). When a model includes time as a fixed effect, time is recalculated to range from 0 at the beginning to 1 at the end of the time window. Thus, the intercept represents model estimates at the onset of the time window, while the estimate for time reflects how much this needs to be adjusted to obtain the estimate at the end of the time-window. In order to obtain the estimates for other factors at the end of the time-window under investigation, time is recentered to range from -1 to 0, prior to reestimating the same model. As these steps do not affect model fit, they are not explicitly reported. In the text, the relevant estimates derived from these models are reported in percentages. When relevant, *p*-values obtained from models based on recentered data are reported in the text, as these cannot be read from the tables.



**FIGURE 4 | Proportion of fixations toward the target image as a function of time for no competitor trials.** Proportions were calculated as subject means over 100 ms bins. Note that Distractor1 is the distractor that has the same gender as the target. Thus, the increased fixation proportion toward this distractor in the late time window reflects rhyme competition (as gender information from the suffix becomes available).



**FIGURE 5 | Proportion of fixations toward the target image as a function of time for target competitor trials.** Proportions were calculated as subject means over 100 ms bins. Here, competition effects can be observed until disambiguating information from the suffix becomes available.

### 3.3.2. Trials Without a Competitor

#### 3.3.2.1. Wordstems

The best fitting model for the time window corresponding to processing the wordstem (200–600 ms after the onset of the target word) contains fixed effects of time and probability along their interaction term (Table 2). Figure 6 shows the estimates obtained from this model relative to the aggregated

participant means. At the beginning of the time-window, the proportion of fixations for low probability targets is estimated at 16.1% (95% CI [11.6, 22.0]). This estimate is not significantly different for medium probability targets (15.9%, 95% CI [11.9, 20.9]), nor for high probability targets (18.4%, 95% CI [12.9, 25.6]). At the end of the time window, the fixation proportion toward low probability targets is estimated at 28.0% (95% CI [21.1, 36.1]). By comparison, medium probability targets have a fixation proportion of 27.7% (95% CI [21.6, 34.7]), which is not significantly different ( $p = 0.943$ ). For high probability items, the estimate is 23.4% (95% CI [16.7, 31.7]), which is also not significantly different ( $p = 0.261$ ).

Although the significant interaction suggests experience-based probabilities to be effective during online processing of the stem, its interpretation is not straight forward. While the main effect of time shows that fixations toward the target increase within this timewindow, the interaction shows that this effect is attenuated for high probability targets. However, since no main effects of probability were detected at the beginning nor at the end of the time-window, the interaction might be spurious, or based on the small but insignificant advantage for high probability targets observed at the beginning of the time-window which is leveled out over time.

#### 3.3.2.2. Suffixes

The best-fitting model for the time window corresponding to processing the suffix (600–1000 ms after the onset of the target word, see Figure 6) contains only a fixed effect of time. This indicates that fixations toward the target increase from 33.2% (95% CI [27.4,39.7]) at the beginning of the time-window to 59.3% (95% CI [52.9,65.3]) at the end of the time-window and is not discussed further here.

### 3.3.3. Trials Featuring a Target Competitor

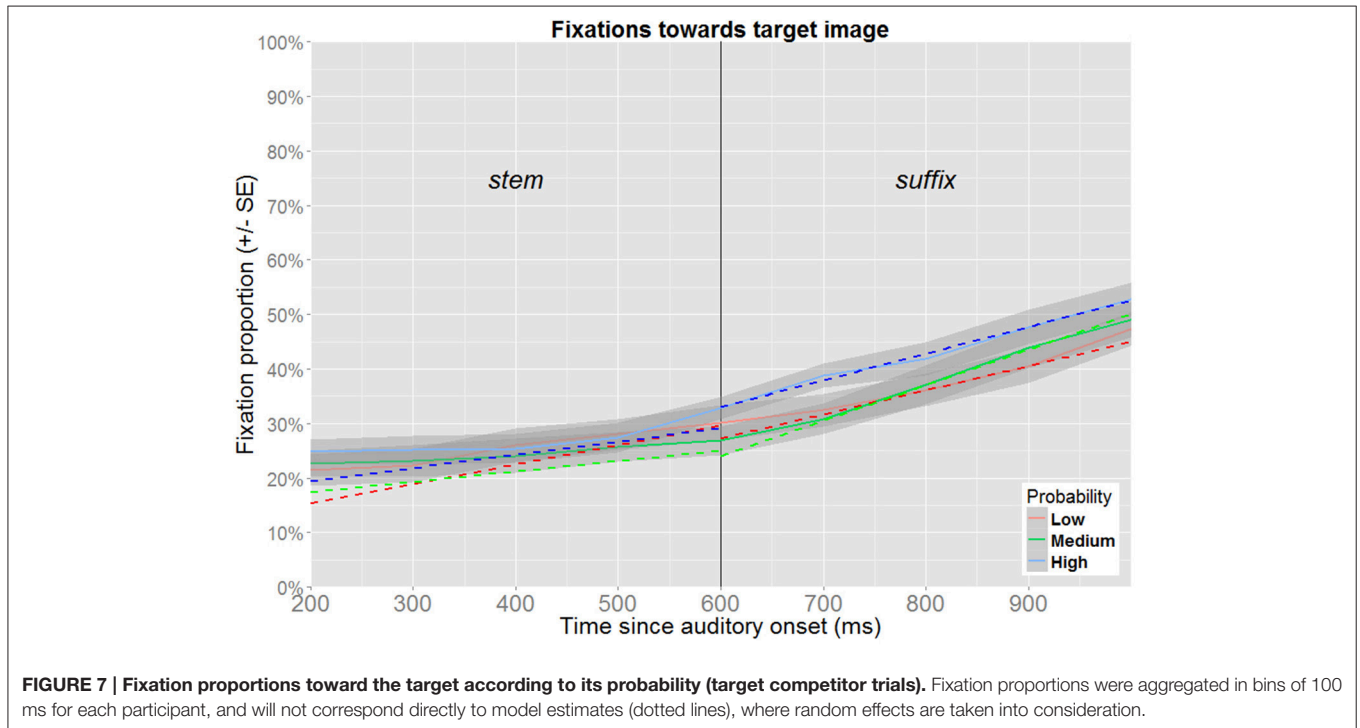
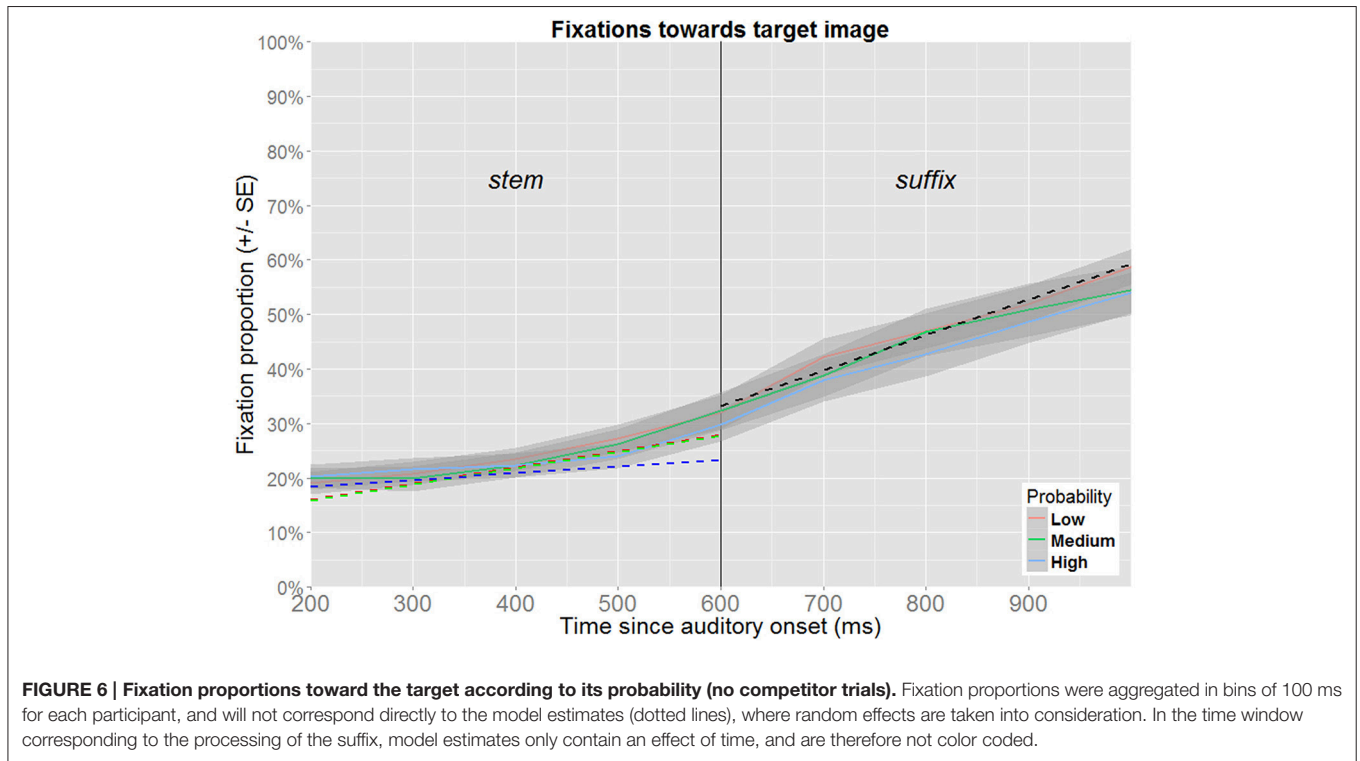
#### 3.3.3.1. Wordstems

For the time-window ranging from 200 to 600 ms after onset of the target word, the best fitting model contains fixed effects for time and probability, along their interaction (Table 3 and Figure 7). At the beginning of the time-window, low probability targets received 15.3% of the fixations (95% CI [8.8,25.4]). Medium probability targets received more fixations (17.4%, 95% CI [13.5,22.1]), but this difference was not significant. The same pattern was found for high probability targets (19.5%, 95% CI [15.3,24.5]), and this is marginally significant ( $p = 0.051$ ). At the end of the time-window, the estimate for low probability targets is 29.7% (95% CI [24.0,36.1]). This is not significantly different from the estimates for medium frequency targets (25.1%, 95% CI [20.1,30.9],  $p = 0.119$ ), nor for high probability targets (29.2%, 95% CI [23.7,35.4],  $p = 0.885$ ).

#### 3.3.3.2. Suffixes

Also in the time window that coincides with the processing of the suffix (600–1000 ms after the onset), the best-fitting model contains fixed effects of time and probability and their interaction (Table 4 and Figure 7). At the beginning of the time-window, low probability targets received 27.3% of the fixations (95% CI [22.7,32.4]). Medium probability targets received less





fixations (24.0%, 95% CI [18.1,31.1]), but this difference was not significant. In contrast, the estimate for high probability targets is significantly higher (33.0%, 95% CI [28.0,38.3]). At the end of the time-window, the estimate for low probability targets is 44.9% (95% CI [37.9,52.2]). This is higher both for medium

probability targets (50.0%, 95% CI [41.2,58.7],  $p = 0.263$ ) and for high probability targets (52.6%, 95% CI [46.8,58.4]), but only the latter contrast is significant ( $p < 0.01$ ).

In summary, the patterns observed in the preceding analyses of eyetracking data indicate that experience-based probabilities

**TABLE 2 | Model estimates, processing word stem (no competitor present).**

	Estimate	SE	z-value	Pr(>t)
(Intercept)	-1.648	0.193	-8.52	<0.001
<b>PROBABILITY</b>				
Medium	-0.016	0.171	-0.09	0.927
High	0.158	0.214	0.74	0.461
<b>TIME</b>				
Time*Medium	0.700	0.171	4.10	<0.001
Time*High	0.004	0.136	0.03	0.979
Time*High	-0.398	0.137	-2.91	<0.01

**TABLE 3 | Model estimates, processing word stem (target competitor present).**

	Estimate	SE	z-value	Pr(>t)
(Intercept)	-1.709	0.321	-5.33	<0.001
<b>PROBABILITY</b>				
Medium	0.153	0.152	1.01	0.314
High	0.292	0.149	1.95	0.051
<b>TIME</b>				
Time*Medium	0.846	0.242	3.50	<0.001
Time*Medium	-0.383	0.135	-2.84	<0.01
Time*High	-0.313	0.133	-2.36	<0.05

**TABLE 4 | Model estimates, processing suffix (target competitor present).**

	Estimate	SE	z-value	Pr(>t)
(Intercept)	-0.981	0.125	-7.87	<0.001
<b>PROBABILITY</b>				
Medium	-0.171	0.181	-0.95	0.345
High	0.271	0.120	2.27	<0.05
<b>TIME</b>				
Time*Medium	0.778	0.160	4.86	<0.001
Time*Medium	0.372	0.122	3.06	<0.01
Time*High	0.037	0.118	0.32	0.753

affect online processing, at least under certain conditions. While no coherent pattern could be detected in the analyses of trials without a competitor, an advantage for high probability targets was detected in the analyses of trials featuring a target competitor. This pattern appeared as a tendency during processing of the stem, and it emerged as a clear effect during the processing of the suffix. While the difference observed between the trial types suggests that the effect is dependent on the available visual information which might have reinforced the relevance of gender information, another plausible scenario would be that the presence of the competitor lead to increased focus on the image pair as such, which made the difference easier to detect. Alternatively, the difference may be due to the overall number of target fixations being higher in the latter trial type, since this also implies that more relevant data were available in the latter analyses.

## 4. GENERAL DISCUSSION

The present study examined experience-based probabilities in the processing of newly acquired word-referent associations. The results show that in the formation of new representations, relative frequencies of exposure to male vs. female versions of the same word/image pairs are tracked, and that these result in expectations about the referent which become active during lexical processing. Methodologically, the study shows that combining artificial language learning with visual world eyetracking (Magnuson et al., 2003), can be successfully applied to investigate the acquisition and online processing of word-reference associations that possess both internal and referential structure.

Processing differences were observed between highly overlapping representations, where male and female referents shared the salient features that were necessary for their identification (color, shape, and texture), and whose distinguishing features (eyes and mouths) were far more subtle. This contrast in salience between shared and distinguishing features holds not only for the visual referents, but also for the pseudowords they were linked to. More specifically, auditory gender cues were presented in word-final position, and were therefore only available after information from the stem had triggered the recognition process. This latter point is verified in the overall gaze patterns, as fixations toward the target image are above chance during online processing (Figures 4, 5). Probabilistic gender cues were redundant during training, and both stems and suffixes occurred equally often in the full stimulus set. That the relative frequencies of exposure nevertheless affected processing in the post-test is striking, since the explicit instructions at the beginning of the experiment highlighted the importance of the linguistically encoded information. As evidenced in the pretest, participants actively and consistently used this information early on.

Response time data show that processing a representation that is consistent with experience-based expectations is faster than processing one that is inconsistent with such expectations. This result is in line with findings for processing of role nouns (as discussed in Section 1.2), and suggests that stereotype effects observed with natural language stimuli can at least in part be attributed to relatively simple aspects of processing (i.e., frequencies of exposure and relatively shallow semantic differences, as simulated in the present study), and do not necessarily depend on the activation of more elaborate semantic information, neither in terms of the complexity of a given representation in itself, or in terms of activating more elaborate contextual information. Response facilitation for frequently presented stimuli is not a novel finding *per se* (e.g., Forster and Chambers, 1973), and could be taken to reflect efficiency of training of exemplars rather than expectations about a category. However, such effects have to our knowledge not previously been observed for highly overlapping, compositionally structured representations, as investigated in the present study. The results presented here suggest that gender-coded words are not treated analytically to the extent that the referential component that is shared by male and female referents can be activated in the

absence of gender information. Had this been the case, we would have expected response times to be independent of the induced expectation.

That gender information available from the suffix was not sufficient to mask the effects of experience-based expectations echoes a finding reported for stereotypical information associated with gender unmarked role nouns: suppressing stereotypical information may be difficult or even impossible (Oakhill et al., 2005). Oakhill et al. instructed participants not to rely on stereotypical information (e.g., when judging whether a “*a sister*” could be “*a plumber*”) and found participants unable to do so. Even in cases where gender information is fully unspecified from the available linguistic information, it does not necessarily follow that this also holds for the mental representation accessed or constructed during language comprehension. An alternative view (e.g., Barsalou, 1999) is that the underlying representations are not abstractions that directly match the associated linguistic structures, but that they by default are highly specified in terms of perceptual detail, as demonstrated in the experiments reviewed in Section 1.1.

Eyetracking data provide additional evidence that effects of experience-based probabilities emerge during online processing of the words, and replicate findings from studies using pseudowords without morphological structure (Magnuson et al., 2003) and natural words without semantic overlap (Dahan et al., 2001). The observation of probability effects during online processing of the pseudowords speaks directly against the possibility that experience-based probabilistic information is inferred only during later processing stages, after a more abstract representation has been retrieved or deeper semantic processing has occurred. When a target competitor was present, participants were more likely to fixate the image whose gender was consistent with the induced expectation than the image whose gender was inconsistent with the induced expectation. This pattern emerged as a tendency during the processing of the stem, and became more robust during the processing of the suffix. Interestingly, this pattern was not detected in the absence of a target competitor. While in principle, this could suggest that access to a direct comparison was necessary for the effect to be detectable, the results from the response times suggest otherwise. That in the absence of a target competitor, low, medium and high probability targets received the same amount of fixations in the timewindows corresponding to the processing of the wordstem and of the suffix also speaks directly against the theoretical possibility that male and female versions of the images were processed as fully independent referents. If participants had acquired one-to-one mappings between unanalyzed words and referents, rather than treating them as overlapping representations, we would have expected the recognition of low-probability targets to be delayed.

We acknowledge that the contrast between the clear patterns observed in the response time data and the weaker (but similar) patterns in the gaze data may not only be due to differences in the sensitivity of these measures, but also to the two measures capturing different stages of processing. Hence, the effects detected in the response time data could be at least partially driven by additional inferences drawn after initial recognition had taken place. Nevertheless, even if the effects of probability

observed in the gaze data were weaker, they clearly show that expectations are triggered during online lexical processing. Crucially, word-referent associations were formed in the absence of contextual information that could have provided additional or indirect cues to a referent's gender, and consequentially the scope of additional inferences would be limited to the established links between words and referents.

In the artificial language used here, suffixation was used as a means to simulate a gender coding system, and the findings reported provide insights complimentary to findings based on studies using natural language materials. Better understanding the activation of gender information during comprehension is relevant for the evaluation of policies to promote gender-fair language use across languages (see e.g., <http://www.unifr.ch/psycho/itn-lcg/en>). In this respect, the reported findings suggest that gender-biases resulting from experience are activated during online processing at the lexical level as opposed to being inferred after a more abstract representation has been retrieved. This would explain why experience-based biases are difficult to overcome. However, inferences about the longevity of the observed effects (e.g., whether they would also affect sentence or discourse level processing) can not be drawn from the available data. Other aspects that would require further research, is how representations may be affected by situations in which gender is unknown or irrelevant, and more specifically how different policies to achieve gender-fair language (e.g., neutralization strategies vs. attempts at enhancing gender visibility) may affect cognitive representation.

The results are expected to generalize to other categories and classifications as well, as the induced expectations were based on probabilities directly derived from experience, and the gender distinction investigated here is just one among many similar distinctions we are confronted with. A plausible account would be that during comprehension, auditory information is used incrementally to identify the associated referent, with the consequence that a word stem is sufficient to activate experimental traces stored in memory (Barsalou, 1999). These memory traces go beyond the information that is available from the linguistic input, to the extent that the features of a referent that has been encountered frequently receive a stronger activation than features of a referent that has been encountered less frequently. As a consequence, even if certain features are not of primary interest or of particular relevance for referential resolution, these features may still play an active role in the (re)construction of a mental representation.

## 5. CONCLUSION

By addressing the relationship between stereotype-based expectations associated with natural language on the one hand and frequencies of exposure in a miniature artificial language on the other, the research presented here compliments literature on cognitive processing of gender information by demonstrating that experience-based processing asymmetries can emerge relatively quickly in the acquisition of a simplified system that does not require deep semantic processing. The results from

the present study suggest (a) that experience-based expectations develop automatically as a consequence of a word's likelihood of being used with reference to either a male or a female referent, and (b) that such expectations are not masked by disambiguating linguistic information. The latter finding indicates that gender information must have been available already during processing of the word stem, and it is difficult to imagine a scenario where this information would be retrieved after disambiguation information has become available. Both findings are consistent with the view that perceptually based information is activated during language comprehension (as outlined in Section 1.1.)

## AUTHOR CONTRIBUTIONS

Both authors contributed extensively to the work presented in this paper. AÖ and DB jointly conceived of the study and sketched the design. AÖ carried out much of the theoretical and practical implementation of the project, and drafted the

full paper. DB supervised all stages of the project. Both authors discussed the results and implications and contributed to the manuscript at all stages.

## ACKNOWLEDGMENTS

The present research was conducted within the Marie Curie Initial Training Network: Language, Cognition, and Gender, ITN LCG, funded by the European Community's Seventh Framework Program (FP7/2007-2013) under Grant Agreement No. 237907 and from the Research Council of Norway under project number 210213.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01250>

## REFERENCES

- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511801686
- Banaji, M. R., and Hardin, C. D. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x
- 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
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/s0140525x99002149
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models using Eigen and S4*. R package version 1.0-6.
- Blender Foundation (2012). *Blender - a 3D Modelling and Rendering Package*. Amsterdam: Blender Online Community, Blender Institute.
- Boersma, P. (2001). Praat, A System for Doing Phonetics by Computer. *Glott Int.* 5, 341–345.
- Corbett, G. G. (1991). *Gender*. Cambridge textbooks in linguistics. New York, NY: Cambridge University Press.
- Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cogn. Psychol.* 42, 317–367. doi: 10.1006/cogp.2001.0750
- Forster, K. I., and Chambers, S. M. (1973). Lexical access and naming time. *J. Verbal Learn. Verbal Behav.* 12, 627–635. doi: 10.1016/S0022-5371(73)80042-8
- Garnham, A., Oakhill, J., and Reynolds, D. (2002). Are inferences from stereotyped role names to characters gender made elaboratively? *Mem. Cogn.* 30, 439–446. doi: 10.3758/BF03194944
- Gómez, R. L., and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4, 178–186. doi: 10.1016/S1364-6613(00)01467-4
- Huetting, F., and Altmann, G. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition* 96, B23–B32. doi: 10.1016/j.cognition.2004.10.003
- Huetting, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Hulstijn, J. H. (1997). Second language acquisition research in the laboratory. *Stud. Second Lang. Acquisit.* 19, 131–143. doi: 10.1017/S0272263197002015
- Kristoffersen, G. (2000). *The Phonology of Norwegian*. New York, NY: Oxford University Press.
- Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. (2014). *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer objects of lme4 package)*. R package version 2.0-6.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., and Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *J. Exp. Psychol. Gen.* 132:202. doi: 10.1037/0096-3445.132.2.202
- Matin, E., Shao, K. C., and Boff, K. R. (1993). Saccadic overhead: information-processing time with and without saccades. *Percept. Psychophys.* 53, 372–380. doi: 10.3758/BF03206780
- Misersky, J., Gygas, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., et al. (2013). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behav. Res. Methods* 46, 1–31. doi: 10.3758/s13428-013-0409-z
- Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cogn.* 33, 972–983. doi: 10.3758/BF03193206
- Pecher, D., van Dantzig, S., and Schifferstein, H. N. J. (2009a). Concepts are not represented by conscious imagery. *Psychon. Bull. Rev.* 16, 914–919. doi: 10.3758/PBR.16.5.914
- Pecher, D., van Dantzig, S., Zwaan, R. A., and Zeelenberg, R. (2009b). Language comprehenders retain implied shape and orientation of objects. *Q. J. Exp. Psychol.* 62, 1108–1114. doi: 10.1080/17470210802633255
- Pyykkönen, P., Hyönä, J., and Van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: evidence from visual world eye tracking. *Exp. Psychol.* 57, 126–133. doi: 10.1027/1618-3169/a000016
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reali, C., Esaulova, Y., Öttl, A., and Von Stockhausen, L. (2015). Role descriptions induce gender mismatch effects in eye movements during reading. *Front. Psychol.* 6:1607. doi: 10.3389/fpsyg.2015.01607
- Reynolds, D. J., Garnham, A., and Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Q. J. Exp. Psychol.* 59, 886–903. doi: 10.1080/02724980543000088
- Richter, T., and Zwaan, R. A. (2010). Integration of perceptual information in word access. *Q. J. Exp. Psychol.* 63, 81–107. doi: 10.1080/17470210902829563
- Richter, T., Zwaan, R. A., and Hoever, I. (2009). Acquiring experiential traces in word-referent learning. *Mem. Cogn.* 37, 1187–1196. doi: 10.3758/MC.37.8.1187

- Rosch, E. (1978). "Principles of categorization," in *Cognition and Categorization*, eds E. Rosch and B. B. Lloyd (Hillsdale, NJ: Lawrence Erlbaum Associates), 27–48.
- Salverda, A. P., Dahan, D., and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90, 51–89. doi: 10.1016/S0010-0277(03)00139-2
- Sato, M., Schafer, A. J., and Bergen, B. K. (2013). One word at a time: mental representations of object shape change incrementally during sentence processing. *Lang. Cogn.* 5, 345–373. doi: 10.1515/langcog-2013-0022
- Singer, J. D., and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. London: Oxford University Press.
- Vukovic, N., and Williams, J. N. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychol.* 145, 98–103. doi: 10.1016/j.actpsy.2013.11.002
- Zwaan, R. A. (2004). The immersed experienter: toward an embodied theory of language comprehension. *Psychol. Learn. Motiv.* 44, 35–62. doi: 10.1016/S0079-7421(03)44002-4
- Zwaan, R. A., and Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PLoS ONE* 7:e51382. doi: 10.1371/journal.pone.0051382
- Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychol. Sci.* 13, 168–171. doi: 10.1111/1467-9280.00430
- Conflict of Interest Statement:** 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.
- The reviewer RM and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.
- The reviewer VH declared a shared affiliation, though no other collaboration, with the authors AÖ and DB to the handling Editor, who ensured that the process nevertheless met the standards of a fair and objective review.
- Copyright © 2016 Öttl and Behne. 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) or licensor 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.*



# How Orthography Modulates Morphological Priming: Subliminal Kanji Activation in Japanese

Yoko Nakano<sup>1,2</sup>, Yu Ikemoto<sup>2</sup>, Gunnar Jacob<sup>3</sup> and Harald Clahsen<sup>3\*</sup>

<sup>1</sup> Graduate School of Language, Communication and Culture, Kwansei Gakuin University, Nishinomiya City, Hyogo, Japan,

<sup>2</sup> School of Human Welfare Studies, Kwansei Gakuin University, Nishinomiya City, Hyogo, Japan,

<sup>3</sup> Potsdam Research Institute for Multilingualism, University of Potsdam, Potsdam, Germany

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Sendy Caffarra,  
Basque Center on Cognition, Brain  
and Language, Spain  
Mark Aronoff,  
Stony Brook University, USA

### \*Correspondence:

Harald Clahsen  
clahsen@uni-potsdam.de

### Specialty section:

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

**Received:** 30 October 2015

**Accepted:** 18 February 2016

**Published:** 30 March 2016

### Citation:

Nakano Y, Ikemoto Y, Jacob G and  
Clahsen H (2016) How Orthography  
Modulates Morphological Priming:  
Subliminal Kanji Activation in  
Japanese. *Front. Psychol.* 7:316.  
doi: 10.3389/fpsyg.2016.00316

The current study investigates to what extent masked morphological priming is modulated by language-particular properties, specifically by its writing system. We present results from two masked priming experiments investigating the processing of complex Japanese words written in less common (moraic) scripts. In Experiment 1, participants performed lexical decisions on target verbs; these were preceded by primes which were either (i) a past-tense form of the same verb, (ii) a stem-related form with the epenthetic vowel *-i*, (iii) a semantically-related form, and (iv) a phonologically-related form. Significant priming effects were obtained for prime types (i), (ii), and (iii), but not for (iv). This pattern of results differs from previous findings on languages with alphabetic scripts, which found reliable masked priming effects for morphologically related prime/target pairs of type (i), but not for non-affixal and semantically-related primes of types (ii), and (iii). In Experiment 2, we measured priming effects for prime/target pairs which are neither morphologically, semantically, phonologically nor - as presented in their moraic scripts—orthographically related, but which—in their commonly written form—share the same kanji, which are logograms adopted from Chinese. The results showed a significant priming effect, with faster lexical-decision times for kanji-related prime/target pairs relative to unrelated ones. We conclude that affix-stripping is insufficient to account for masked morphological priming effects across languages, but that language-particular properties (in the case of Japanese, the writing system) affect the processing of (morphologically) complex words.

**Keywords:** morphologically complex words, morpho-orthography, decomposition, Japanese, kanji, kana

## INTRODUCTION

The processing of morphologically-complex words has been subject to considerable debate in the past two decades. A core question in this area of research concerns the mechanisms the processing system employs for morphologically-complex words during word recognition. A number of studies have used the masked priming technique to examine this question. In a masked priming experiment, prime words are presented for a very short period of time only, which typically prevents the prime words from being directly recognized. Instead, masked priming is supposed to tap into subliminal processes involved in visual word recognition. A considerable number of studies have shown that native speakers can extract morphological information from inflected and derived words under masked priming conditions, by showing masked priming

effects for morphologically complex word forms independently of the activation of semantic information and beyond pure orthographic priming; see Marslen-Wilson (2007). Masked morphological priming effects are supposed to be due to a morpho-orthographic segmentation mechanism that identifies the word root by stripping off affixes at an early stage of processing. Evidence for this affix-stripping mechanism comes from a series of masked priming experiments in different languages, including English (Rastle et al., 2000, 2004; Silva and Clahsen, 2008), French (Longtin and Meunier, 2005), Arabic (Boudelaa and Marslen-Wilson, 2005), Russian (Kazanina et al., 2008), German (Neubauer and Clahsen, 2009; Clahsen and Neubauer, 2010), Basque and Spanish (Duñabeitia et al., 2007). For example, Rastle et al. (2004) found that pseudo-affixed primes (e.g., *brother*), which consist of a potential but non-existent stem + affix combination (*broth-er*), prime semantically unrelated targets (e.g., *broth*) as efficiently as derivationally related prime-target pairs (e.g., *cleaner—clean*) in the masked priming paradigm. At the same time, however, primes that contain non-affixal segments (e.g., *brothel*) did not produce a corresponding (pseudo)-stem priming effect. These contrasts provide support for a decomposition mechanism by which affixes are automatically stripped off their stems, even for semantically unrelated prime-target pairs such as *brother/broth*. The presence of a pseudo-affix (e.g., *-er* in *brother*) is apparently sufficient to trigger affix stripping. Similar contrasts have also been reported for French and Russian (Longtin and Meunier, 2005; Kazanina et al., 2008).

Note, however, that all these studies come from languages with alphabetic scripts in which morphological and orthographic boundaries typically coincide. To take an example, Berg and Aronoff (submitted) demonstrate that the spelling of both inflected and derived words in English marks morphological information and that homography of suffixes and homophonous word endings tend to be avoided. For instance, while English words that end in the letters <ous> are denominal adjectives with <ous> corresponding to the suffix that derives adjectives from nouns, the same phonological sequence [əʊ] is spelled differently when it is not an adjective; compare, for example, [nɜːvəs]<sub>Adjective</sub> with [sɜːvɪs]<sub>Noun</sub>. Berg and Aronoff (submitted) show that these grapheme/morphology contingencies are not accidental, and that English spelling is not only lexically but also morphologically determined. Consequently, given the properties of this kind of writing system, affix stripping appears to be a particularly sensible strategy for word reading.

This then raises the question of whether affix stripping under masked priming conditions universally applies in morphological parsing, irrespective of a language's particular writing system. This question is still open as there are only very few studies to date that have examined masked morphological priming in languages with non-alphabetic writing systems (but see Clahsen and Ikemoto, 2012; Fiorentino et al., 2015). Against this background, the present study reports results from masked priming experiments in Japanese, a language in which morphological segmentation and orthographic boundaries sometimes fail to coincide.

## BACKGROUND: SOME BRIEF NOTES ON JAPANESE ORTHOGRAPHY

There is a vast amount of literature on the different writing systems of Japanese, a discussion of which goes beyond the scope of the current study; see Sampson (1985, Chapter 9) for a review. Instead, this section presents a brief descriptive overview for those unfamiliar with Japanese.

Japanese has a mixed writing system consisting of *kanji*, the logographs adopted from Chinese, and *kana*, a syllable—more specifically mora-based phonographic writing system. Like the Chinese logographs, the Japanese *kanji* are associated with particular meanings. However, unlike the Chinese graphs (which basically have one pronunciation), *kanji* often allow distinct ways of pronunciation in Japanese. *Kana* scripts, on the other hand, are orthographic signs (“syllabograms,” Coltheart, 2014) for particular sounds, with each *kana* typically encoding a particular combination of a consonant and a vowel. The *kana* script comes in two subtypes: hiragana and katakana.

The mixed Japanese orthographical system allows users to choose whether to write a word in *kanji* plus *kana* or in *kana* only. Typically, words of Chinese origin and the meaningful parts of native Japanese words (i.e., the roots) are written with *kanji*, while grammatical morphemes (e.g., inflectional affixes) and additional elements added to the root appear in hiragana. In addition, loanwords from European languages and foreign names are entirely written with katakana. Furthermore, due to the phonographic nature of *kana*, it is possible (though uncommon) to write any Japanese word with *kana* only, even those that are normally written with the mixed script. Following Saito (1981), several experimental studies have investigated processes involved in lexical access from the different scripts; see discussion in Dehaene (2009) and Coltheart (2014). Due to its logographic nature, reading *kanji* engages what is labeled the “lexical” route in dual-route reading models (Coltheart et al., 2001), which provides a link from the orthographic to the semantic lexicon. By contrast, due to its phonographic nature, reading *kana* engages the “non-lexical” route for converting subword-level orthographic units (viz. *kana*) to subword-level phonological units (viz. mora). Support for this contrast comes, for example, from Chen et al.'s (2007) lexical priming experiment in which *kanji* targets (e.g., 展示 *tenji* “display”) were presented at short (85 ms) and long (150 ms) intervals preceded by three types of prime: (i) homophonic primes (e.g., 点字 *tenji* “braille”), (ii) semantically related primes (e.g., 陳列 *chinretu* “display”), and unrelated ones (e.g., 流浪 *ruro* “wandering”). When the prime words were presented with *kanji*, only the semantically related condition produced priming effects at both intervals. However, when the prime words were presented with hiragana only, the homophonic condition showed priming effects at both intervals, and the semantically related condition showed a priming effect only at the long interval. These results indicate that the lexical route is used for reading *kanji* and the non-lexical one for reading *kana*.

Regarding the spelling of morphologically complex words, it is important to note that orthographic boundaries do not always coincide with morphological segmentation in Japanese. Instead,

the root-final phoneme may form a new mora with an affix or other segment, which is then spelled with a kana. Consider, for example, the consonant-final stem 眠 *r* /nemur/ “sleep.” When inflectional endings are added to this stem, e.g., the imperative *-e* or the past tense *-ta*, the stem-final consonant forms a mora with the suffix, which is spelled with kana. The imperative form, for example, is 眠れ “sleep!” with the kana れ for /re/. The past-tense form is 眠った “slept” with the stem-final /r/ changed to /t/ and the /tta/ segment spelled with the two kana つ and た.

## THE PRESENT STUDY

The current study examines whether morpho-orthographic decomposition (“affix-stripping”) of morphologically complex words, as reported in masked morphological priming studies of languages with alphabetic scripts, is also employed in a language with a different writing system (viz. Japanese). In previous research, priming effects under masked stimulus conditions have been reported for morphologically complex words in Japanese (Clahsen and Ikemoto, 2012; Fiorentino et al., 2015), specifically for deadjectival forms with the suffixes *-sa* and *-mi*. In both studies, significant masked priming effects were found irrespective of whether primes and targets were spelled in the common mixed script (i.e., with kanji plus kana) or in kana only. While the priming effects might indeed—as suggested by the authors of the two studies—be due to affix stripping (of *-sa* and *-mi*), parallel to the masked priming effects found for derived (and inflected) words in English and other languages with alphabetic scripts, an alternative possibility needs to be considered. This is because the critical prime-target pairs that yielded significant facilitation of target recognition times in Clahsen and Ikemoto’s and in Fiorentino et al.’s studies—when written with the common mixed script—share the same kanji. Consequently it is conceivable that the reported priming effects are due to this overlap in kanji, rather than (or perhaps on top of) affix stripping. Consider, for example, the prime-target pair 楽しい - 楽しみ “delightful-delightfulness” from Fiorentino et al. (2015, **Table 1**) from which the orthographic overlap in terms of shared kanji is obvious. Clahsen and Ikemoto (2012) tried to reduce this kind of direct orthographic overlap between primes and targets by presenting their stimuli in the kana-only script, with the primes and the targets in hiragana. If, however, these stimuli are written in the normal (mixed) script, the prime-target pairs that yielded priming effects in their experiment share the same kanji and the ones that do not show priming have different kanji. Indeed, Clahsen and Ikemoto (2012) found the same priming patterns in a follow-up experiment in which all items were presented in the mixed script (with kanji) as in their main experiment with kana-only stimulus presentation. Given these findings, it is conceivable that activation of the mixed script including the corresponding kanji cannot be completely blocked, even when reading the prime-target pairs in the unusual kana-only script.

To further elucidate the nature of masked priming effects of complex words in written Japanese, the present study addresses two questions. Experiment 1 asks whether decomposition of complex words under masked priming conditions is genuinely morphological (viz. “affix stripping”) or whether non-affixal

material is also segmented from the root in Japanese. Experiment 2 asks how masked priming effects in Japanese are modulated by the particular properties of its mixed writing system, specifically how the activation of kanji affects visual word recognition.

## EXPERIMENT 1: AFFIX STRIPPING IN JAPANESE MASKED PRIMING

Affix stripping is a powerful morphological parsing mechanism that the masked priming technique is supposed to tap into. Stanners et al. (1979) explained morphological priming effects as follows: “...the base verb and suffix are partitioned prior to memory access and the base verb is then directly accessed” (p. 403). In other words, when a word form such as *walked* is presented as a prime, the affix *-ed* is stripped off, thereby isolating the base stem which then directly facilitates recognition of a target word such as *walk*. Crucially, non-affixal segments of morphologically unrelated words have been shown not to produce masked priming effects in English and other languages with alphabetic scripts; see Marslen-Wilson (2007). In English, for example, *darkness* primes *dark*, but *example* does not prime *exam*, reflecting the fact that (unlike <ness>) the letters <ple> do not function as an affix (Heyer and Clahsen, 2015).

In the present masked priming experiment, we examined whether this contrast also applies to Japanese. Two types of critical prime words were tested: (i) inflected *-ta* suffixed past-tense verb forms and (ii) words with the non-affixal word-final segment /i/. Both critical types of prime words have parallel surface forms with one segment added to the stem. However, while prime type (i) is a morphologically structured word form with the past-tense suffix *-ta*, prime type (ii) is the non-affixal infinitive form of consonant-final verbal stems, the most common type of verbal stem in Japanese. Crucially, the /i/ segment added to these verbal stems represents a case of phonological epenthesis (enforced by the CV phonotactics of Japanese), rather than an inflectional or derivational suffix (Kiyose, 1971; Shirota, 1998; Tagawa, 2012); see examples (1) and (2) below. The target forms used for both prime types were non-past forms of the same verbs, which consist of the same stems as the primes plus the invariant non-past affix *-u*; see example (3). With these conditions we can directly compare priming from affixed vs. non-affixal word forms on the same target words. Two additional conditions were added to assess the potential contribution of semantic and phonological relatedness. The first condition consisted of primes and targets that were semantically related. They were either synonyms (e.g., ほめる *homeru* “compliment”—タタエル *tataeru* “praise”); see examples (4) and (5) or semantic associates (e.g., あるく *aruku* “walk”—ハシル *hashiru* “run”); see examples (6) and (7). The second condition constituted primes and targets that were phonologically related, i.e., similar-sounding but otherwise unrelated prime-target pairs, e.g., たたむ *tatamu* “fold”—タタカウ *tatakau* “fight”; see examples (8) and (9). Assuming that under masked priming conditions affixed word forms are morphologically decomposed (“affix stripping”) in Japanese, we would expect to find a reliable morphological priming effect for *-ta* forms, i.e., prime type (i), but not for the non-affixal /i/ forms of prime type (ii),



parallel to the masked priming results reported for English and other languages with alphabetic scripts. Furthermore, under the assumption (e.g., Rastle et al., 2000, 2004) that masked priming effects are due to morpho-orthographic segmentation (viz. affix stripping), we do not expect the semantic and phonological control conditions to yield any reliable priming effects.

(1) Prime type (i): *-ta* form, e.g., *nemutta* “slept” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ne	mu <sup>rβ</sup>	t	ta
Hiragana	ね	む	っ	た
Kanji (root) + Kana	眠		っ	た

(2) Prime type (ii): */i/* verb form, e.g., *nemuri* “sleep” written with kana only and kanji plus kana

	stem		epenthesis	
International Phonetic Alphabet	ne	mu <sup>rβ</sup>	r <sup>J</sup>	i
Hiragana	ね	む		り
Kanji (root) + Kana	眠			り

(3) Target form: non-past verb form, e.g., *nemuru* “sleep” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ne	mu <sup>rβ</sup>	r	u <sup>rβ</sup>
Katakana	ネ	ム		ル
Kanji (root) + Kana	眠			る

(4) Prime type: non-past verb form, e.g., *homeru* “compliment” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ho	me	r	u <sup>rβ</sup>
Hiragana	ほ	め		る
Kanji (root) + Kana	褒	め		る

(5) Target form: non-past verb form, a synonym, e.g., *tataeru* “praise” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ta	ta	e	ru <sup>rβ</sup>
Katakana	タ	タ	エ	ル
Kanji (root) + Kana		讃	え	る

(6) Prime type: non-past verb form, e.g., *aruku* “walk” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	a	ru <sup>rβ</sup>		ku <sup>rβ</sup>
Hiragana	あ	る		く
Kanji (root) + Kana		歩		く

(7) Target form: non-past verb form, a semantic associate, e.g., *hashiru* “run” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ha	ʃi	r	u <sup>rβ</sup>
Katakana	ハ	シ		ル
Kanji (root) + Kana		走		る

(8) Prime type: non-past verb form, e.g., *tatamu* “fold” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ta	ta	m	u <sup>rβ</sup>
Hiragana	た	た		む
Kanji (root) + Kana		畳		む

(9) Target form: phonologically-related non-past verb form, e.g., *hashiru* “run” written with kana only and kanji plus kana

	stem		suffix	
International Phonetic Alphabet	ta	ta	ka	u <sup>rβ</sup>
Katakana	タ	タ	カ	ウ
Kanji (root) + Kana		戦		う

### Participants

Twenty-eight Japanese speakers [mean age: 22 (*SD*: 6.73), age range: 18–45, 16 females and 12 males] were recruited from the undergraduate and graduate communities at Kwansai Gakuin University in Japan. All participants had normal or corrected-to-normal vision. This study was carried out in accordance with the recommendations of the Grant-in-Aids for Scientific Research of the Japan Society for the Promotion of Science, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### Materials

We constructed 24 experimental item sets, with each set consisting of four prime-target pairs. Each set was based on one of 24 Japanese verbs with consonant-final stems, with the non-past form of the respective verb (e.g., *nemuru*) serving as the target word in all four prime-target pairs of the respective item set; see Appendix A in Supplementary Material for the complete set of items. The target was preceded by one of four different primes: (i) the non-affixal */i/* form of the given verb stem, (ii) the corresponding past-tense form with the inflectional suffix *-ta*, (iii) a matched unrelated control prime, or (iv) an identity control prime in which the target word occurred as both prime and target. Priming effects were determined by comparing the mean RTs for the target words following */i/* and *-ta* primes to those following unrelated control primes. The identity control condition was added as a manipulation check of whether participants were sensitive to any properties of the prime words at all under masked presentation conditions. If this is the case, we would expect to find a repetition priming effect for

Identity primes relative to Unrelated primes, i.e., reduced target RTs for the former relative to the latter.

Word-form frequencies of the items for the two prime types (i) and (ii) were matched as closely as possible using Amano and Kondo's (2003) frequency dictionary, which contains more than 340,000 words collected from a Japanese newspaper between 1985 and 1998. The mean word-form frequencies (per million) were 13.6 for the *-ta* and 19.8 for the */i/* forms, a non-significant difference [ $t_{(23)} = 1.624, p = 0.118$ ] (*-ta* vs. */i/*). Unrelated primes were also matched in length (mora) and with respect to word-form frequency to the targets [length:  $t_{(23)} = 1.282, p = 0.213$ , frequency:  $t_{(23)} < 1$ ]. As prime types (i) and (ii) are semantically, orthographically and phonologically related to their target forms [see (1) to (3) above], the potential contributions of these properties need to be considered. To reduce potential effects of orthographic relatedness between primes and targets, we used the two distinct moraic scripts of Japanese, hiragana, and katakana. All prime words were presented in hiragana whereas all targets were presented in katakana; this is illustrated in (1) to (3) for the different prime and target forms of the verbal stem *nemur-*. Two additional conditions were added to assess the potential contribution of semantic and phonological relatedness. In the semantic control condition, there were 24 semantically-related prime-target pairs, 12 consisting of synonyms (e.g., ほめる *homeru* "compliment"-タタエル *tataeru* "praise") and 12 semantic associates (e.g., あるく *aruku* "walk"-ハシル *hashiru* "run"). These item pairs were selected from an offline rating task in which an additional group of 22 native Japanese speakers (none of whom participated in the main experiment) rated the semantic relatedness of 96 word pairs on a 7-point scale. The 24 pairs selected for the main experiment received significantly higher semantic relatedness ratings than the unrelated control pairs [means: 5.1 (*SD*: 0.71) vs. 1.8 (*SD*: 0.7),  $t_1(21) = 14.665, p < 0.001$ ;  $t_2(46) = 15.568, p < 0.001$ ], which confirmed the semantic relatedness of the test items. The second control condition consisted of phonologically related, i.e., similar-sounding but otherwise unrelated prime-target pairs, e.g., たたむ *tatamu* "fold"-タタカウ *tatakau* "fight." There were again 24 item sets in this condition. For the semantic and the phonological control conditions, related and unrelated primes were matched for length (in terms of mora) and for word-form frequency (all  $t_s < 1$ ).

The experimental items from the 72 item sets were distributed across four different presentation lists according to a Latin-square design, with each presentation list containing exactly one prime/target pair from each item set. As a result, each participant saw each target word only once, ensuring that no participant made repeated lexical decisions on the same target word. We added 328 filler items, resulting in a total of 400 trials per presentation list. In order to make the lexical-decision task meaningful, the target words in 200 of the 328 fillers were non-words. Thus, within each presentation list, 200 targets were existing words, and the other 200 were non-words. Non-words were created by changing one or two mora of an existing word. The order of items was pseudo-randomized, ensuring that experimental items did not appear adjacent to each other.

## Procedure, Data Scoring, and Analysis

Each trial started with a fixation point appearing for 500 ms in the middle of the screen, followed by a 500 ms blank screen, after which a forward mask was presented for 500 ms. Prime words were presented immediately after the mask, and remained on screen for 50 ms. At the offset of the prime, the corresponding target word was presented for 1000 ms. The next trial started 500 ms after the response or timeout. Participants were instructed to make a lexical decision to the target word by pressing one of two buttons as quickly and accurately as possible. The experiment started with a practice session with 10 items. During the experiment, three breaks were provided after every 100 trials. The presentation of the stimuli and the measurement of the reaction times were controlled by the DMDX software package (Forster and Forster, 2003). The whole experiment lasted ~20–30 min.

Timeouts (response times above 2500 ms) and trials with incorrect lexical decisions were excluded from further analyses of the reaction time data. These criteria led to the removal of 13.1% of the trials from the */i/* and 9.9% from the *-ta* condition (the numerical difference in exclusion rates between the */i/* and *-ta* conditions was non-significant,  $t_{1/2} < 1$ ), with 16.1% from the semantic overlap and 22.1% from the phonological overlap condition. These exclusion rates are higher than usual for lexical-decision tasks with native speakers, but note that the participants performed lexical decisions on target words written in an unusual script.

In addition, lexical decision times which were more than 2.5 *SDs* above or below the overall participant mean were considered outliers and therefore also removed; this affected 3.4 % of all *-ta* and */i/* trials, 2.1% of all "semantic" overlap trials and 3.0 % of all "phonological" trials.

## Results

Mean lexical decision times and standard deviations by prime type and condition are shown in **Table 1**. Note that because */i/* and *-ta* primes were tested on the same targets, Unrelated and Identity primes had the same RTs in these two conditions.

As a manipulation check of whether participants were able to retrieve any information from the masked hiragana-spelled primes at all, we first tested for repetition-priming effects by comparing target RTs for Identity vs. Unrelated primes. Paired-samples *t*-tests revealed significant repetition priming effects for all three conditions [*/i/-ta* condition:  $t_1(27) = 5.29, p < 0.001$ ;  $t_2(23) = 5.57, p < 0.001$ ; semantic overlap:  $t_1(27) = 4.93, p < 0.001$ ;  $t_2(23) = 5.75, p < 0.001$ ; phonological overlap:  $t_1(27) = 5.40, p < 0.001$ ;  $t_2(23) = 4.43, p < 0.001$ ] suggesting that

**TABLE 1 | Mean lexical decision times (and standard deviations) by condition and prime type.**

	<i>/i/</i> form	<i>-ta</i> form	Semantic overlap	Phonological overlap
Identity	631 (68)		638 (76)	651 (72)
Test	646 (94)	660 (92)	700 (73)	715 (88)
Unrelated	741 (102)		731 (90)	727 (61)

participants were sensitive to properties of the hiragana-spelled prime words despite the fact that they were written in an unusual script and were masked with only a 50 ms presentation time.

The critical test of our hypotheses requires comparisons between the Test and Unrelated conditions. One-way ANOVAs comparing RTs for targets following /i/, *-ta*, and Unrelated primes revealed a significant main effect of Prime Type [ $F_1(2, 54) = 18.37, p < 0.001$ ;  $F_2(2, 46) = 11.34, p < 0.001$ ]. To explore the source of this effect, we determined priming effects separately for /i/ and *-ta* primes. Paired-samples *t*-tests revealed a difference between the Test and Unrelated conditions for both the *-ta* prime [ $t_1(27) = 4.51, p < 0.001$ ;  $t_2(23) = 3.57, p = 0.002$ ] and the /i/ prime conditions [ $t_1(27) = 5.376, p < 0.001$ ;  $t_2(23) = 4.352, p < 0.001$ ]. With regard to the semantic and phonological control conditions, paired samples *t*-tests revealed a significant difference between the Test and Unrelated prime types for the semantic overlap condition [ $t_1(27) = 2.49, p < 0.05$ ;  $t_2(23) = 2.17, p < 0.05$ ], but not for the phonological overlap condition [ $t_1(27) = 1.00, p = 0.326$ ;  $t_2(23) = 0.90, p = 0.379$ ]<sup>1</sup>.

In sum, Experiment 1 revealed significant priming effects for *-ta* forms, /i/ forms, and for the semantic overlap condition, but no priming effect for the phonological overlap condition. In comparison with previous masked priming studies of languages written in alphabetic scripts, this data pattern is unusual in two ways. First, our results show a significant priming effect for semantically related prime-target pairs, suggesting that in Japanese the processor was able to access semantic information from the primes under masked priming conditions, unlike what has been reported in most previous studies on masked morphological priming which have not found any reliable semantic priming effects and have claimed that the particular early stage of processing tapped by masked priming is semantically blind; see Marslen-Wilson (2007) and Davis and Rastle (2010) for a discussion. Second, inflectionally related prime words with the past-tense suffix *-ta* were found to reliably facilitate target recognition, as was found in languages with alphabetic scripts, so were word forms with the non-affixal segment /i/, with a similar magnitude as the *-ta* forms, indicating that morphological decomposition (viz. “affix stripping”) appears to be insufficient to explain the observed priming effects.

Consider a number of alternative possibilities to explain the data pattern obtained in Experiment 1. First, as we found reliable priming in the semantic control condition, it is conceivable that priming in the critical *-ta* and /i/ conditions might also be semantic in nature, since primes and targets are semantically related in these conditions. This would be in line with claims made by Feldman and collaborators who argued that semantic information from complex words can be accessed under masked

priming conditions (e.g., Feldman et al., 2009, 2015). Note, however, that the magnitudes of priming for the two critical conditions were considerably larger than for the semantic control condition (95 and 81 ms for the former vs. 31 ms for the latter; see **Table 1**) indicating that semantic relatedness cannot fully explain the facilitation effect obtained from *-ta* and /i/ forms. Instead, each prime-target pair in the two critical conditions contains two word forms of the same lemma. Hence, lexical identity (after morphological analysis) seems to be the crucial source of the critical priming effects.

A second possibility might be that languages differ in how complex words are segmented. While affix stripping appears to be a powerful mechanism for visual word recognition of complex words in English, for example, processing complex words in Japanese might not rely on affix stripping. Instead, the system may directly search for possible roots when reading complex words in Japanese (“root spotting”). Given this mechanism, the type of segment that a potential root is combined with—whether or not it is an affix—is irrelevant. Since roots are shared between primes and targets in both the *-ta* and the /i/ conditions, root spotting may explain our finding that the *-ta* and the /i/ conditions produced priming effects of a similar magnitude. Root spotting may also account for the priming effect we found for semantically related prime-target pairs, which was, however, reduced relative to the magnitude of the priming effect obtained for the two critical conditions. This contrast could be due to the fact that in the *-ta* and the /i/ conditions, the prime and the target words share the same root, whereas in the semantic control condition the prime and the target words had different but semantically related roots. Note, however, that while a root-spotting mechanism might be operative in reading complex words in Japanese, the masked morphological priming results for English, French, and other languages cannot be explained in these terms. Recall that in English, for example, pseudo-morphologically related prime-target pairs such as *brother-broth* yielded significant masked priming effects, while prime/target pairs such as *brothel-broth* did not, although root spotting would have predicted the root *broth* to be easily identifiable from both primes, *brother* and *brothel*, which should have yielded parallel facilitation effects. This was not the case, however. It is thus possible that root-based decomposition is a language-specific property of Japanese.

A third possibility is based on the particular properties of the Japanese writing system. Recall that primes and targets in Experiment 1 were presented in different kana scripts (primes in hiragana, and targets in katakana), and were thus not directly related, neither visually nor orthographically. If the words had been presented in the mixed script with kanji, however, the corresponding kanji versions of the primes and targets in both the *-ta* and /i/ conditions would have shared the same kanji. For illustration, consider the two primes and the target in (1) to (3) above. When written in the mixed script, both primes (the *-ta* and the /i/ forms) share the same initial Kanji [viz. 眠 in (1) to (3)] with their target word forms. Hence, if Japanese readers do not completely block the (more common) mixed script, even when reading words with kana only, they will activate the shared kanji in both the primes and the target words. This may then

<sup>1</sup>As regards corrections for multiple comparisons, note that although the design included three prime types, the Identity primes only served as a manipulation check to see whether the experimental technique (viz. masked priming) works with our Japanese stimuli and was not critical to test any of our hypotheses. Consequently, the phonological and semantic control conditions only consist of two critical prime types and therefore do not require any multiple comparison correction. On the other hand, the analyses for the morphological item set involve comparisons between three conditions (/i/, *-ta*, and unrelated). However, the differences between these conditions were robust enough even after Bonferroni-correction.

cause the observed facilitation effect for prime words with *-ta* and */i/*. In this way, (indirect) kanji activation might be a third possible source for the observed priming effects. Experiment 2 was designed to further elucidate the sources of masked priming in Japanese.

## EXPERIMENT 2: ACCESSING KANJI IN READING MORAIIC SCRIPT

Experiment 2 was designed to distinguish between the different accounts described above. Specifically, we measured priming effects for prime/target pairs that were presented— as in Experiment 1— with kana only. However, even though the prime/target pairs for Experiment 2 were unrelated in every single aspect they shared kanji if spelled with the mixed script. Consider for illustration the prime-target pair in (4), with (4a) showing both prime and target in kana only (as they appeared on screen in Experiment 2) and (4b) for the corresponding mixed script version:

- (4) a. とおり - ツウ *toori* ‘street’ - *tsuu* ‘expert’  
b. 通り - 通

Given that the critical prime-target pairs in Experiment 2 are unrelated, except for the shared kanji (e.g., 通 in 4b), the results from this experiment should allow us to decide whether or not kanji are activated while reading Japanese words with moraic scripts. If this is the case, we would expect to find facilitated target recognition for prime-target pairs such as (4a), due to the (indirectly activated) shared kanji, relative to unrelated items in which prime and target words do not have a kanji in common. Alternative sources of priming, on the other hand, should not play a role for the prime-target pairs tested in Experiment 2, as these primes and targets are neither morphologically, semantically, orthographically nor phonologically related.

### Participants

A new group of 23 native speakers of Japanese was recruited, none of whom had participated in Experiment 1 [mean age: 31.5 (*SD*: 12.34), age range: 18–57, 15 females and 8 males]. All participants had normal or corrected-to-normal vision.

### Materials

We constructed a total of 16 experimental item sets, with each item set consisting of a Test prime-target pair and an Unrelated control pair. Within each item set, the Test pair consisted of the */i/* form of a consonant-final verbal stem (e.g., *hakobi* “process”) as prime and a monomorphemic word (e.g., *un* “luck”) as target; see Appendix B in Supplementary Material for the complete item set. Primes and targets were morphologically, semantically, and phonologically unrelated, but, when written with the mixed script, share a kanji, e.g., 運 in *hakobi-un* (運び-運). However, as in Experiment 1, stimulus presentation was in kana only, with the primes being presented with hiragana and the targets with katakana (e.g., はこびーウㇿ *hakobi-un*). The Unrelated control pair consisted of the same target word as the Test pair preceded by a matched unrelated control prime (e.g., さわぎ *sawagi* “disturbance”). Unlike the Test condition,

Unrelated prime words (if spelled with the mixed script) did not share any kanji with their respective targets, e.g., *sawagi-un* (騒ぎ-運). To determine kanji-mediated priming effects, we compared RTs to the target words following the Test primes to those following unrelated primes. Word-form frequencies of related and unrelated primes were matched as closely as possible using the jpTenTen corpus [LUW, sample] (long-unit words) in Sketch Engine (<https://www.sketchengine.co.uk/>), which consists of about 163 million words. The mean word-form frequencies (per million) were 11.64 for the Test prime words and 13.84 for the Unrelated prime words; the difference between the two was non-significant [ $t(15) = 0.09, p = 0.93$ ]. Test primes and Unrelated primes were also matched for word-form frequency with the target words. The mean word-form frequency for the targets was 27.52, which was not significantly different from either Test primes or Unrelated primes [Test vs. Target:  $t(15) = 1.295, p = 0.215$ ; Unrelated vs. Target:  $t(15) = 1.494, p = 0.156$ ]. The mean length (in mora) was also matched between Test and Unrelated primes [3.19 vs. 3.31,  $t(15) = 0.565, p = 0.58$ ]. We also calculated the number of shared phonemes between prime and target in the related and the unrelated conditions. The mean phonological overlap was low (0.9 for Test and 1.3 for Unrelated) and not reliably different between conditions [ $t(15) = 2.1, p = 0.11$ ].

The items from the 16 item sets were distributed across two presentation lists according to a Latin-square design, with each list containing exactly one of the two prime-target pairs from each set. We also added a total of 40 filler items, 12 word-word pairs, and 28 word-nonword pairs. The word-word fillers consisted of word forms with the epenthetic word-final segment */i/* (e.g., すくゝ *sukui*, “saving”) and other monomorphemic nouns (e.g., はは *haha*, “mother”). These prime-target pairs were phonologically, morphologically, and semantically unrelated, and had different kanji when spelled in mixed script writing. The non-word filler targets also had */i/* as the final segment, but did not constitute existing words in Japanese.

### Procedure

The experimental procedure was parallel to Experiment 1. Prime words were presented for 50 ms. Incorrect responses and timeouts (2500 ms) were excluded from further analyses (6.5% of all experimental trials), with similar proportions of excluded data for the Test (6.4%) and the Unrelated conditions (6.4 vs. 6.6%,  $t_{1/2} < 1$ ). Lexical decision times which were more than 2.5 *SDs* above or below the overall participant mean were considered outliers and were removed (2.8% of all test trials). In addition, the results from one participant had to be excluded from any further analysis because of a high overall error rate (50% of all critical trials). We also excluded two items because of high error rates across participants (34.8 and 39.1%).

### Results

Mean lexical decision times were 693 ms (*SD*: 76) for the Test condition and 727 ms (*SD*: 113) for the Unrelated condition, a significant difference for both participants and items [ $t_1(21) = 2.406, p = 0.025$ ;  $t_2(13) = 2.591, p = 0.022$ ].

This finding from Experiment 2 demonstrates a kanji-mediated priming effect for stimuli that were entirely written in kana only, and were otherwise unrelated. We conclude from this finding that Japanese readers activate kanji while reading words written in moraic scripts. In the following, we discuss the implications of this finding for the results of Experiment 1 and with respect to masked priming effects for complex words in Japanese more generally.

## GENERAL DISCUSSION

The current study investigated how the processing of complex words in Japanese is modulated by properties of its writing system. Experiment 1 showed significant priming effects of a similar magnitude on the same target words from both past-tense forms with the inflectionally suffix *-ta* and morphologically simplex word forms with a non-affixal (epenthetic) segment (/i/). In Experiment 2, we found a significant priming effect for prime-target pairs which were entirely unrelated except for the fact that prime and target share a kanji, if they had been spelled with the common mixed script. Recall, however, that in order to avoid any direct visual after-image and/or orthographic overlap between prime and target, the stimuli in both experiments were actually not presented with the mixed script, but were entirely written with the moraic scripts (primes in hiragana, and targets in katakana). Nevertheless, the results of Experiment 2 indicate that the processor activates the corresponding mixed-script versions of the stimuli even when processing words entirely presented with kana. Also note that indirect kanji activation is subliminal and automatic, as the primes were presented under masked priming conditions, which prevented participants from consciously recognizing the prime words. A related effect has been obtained by Thierry and Wu (2007) who observed unconscious translation effects in Chinese/English bilinguals' reading of English words.

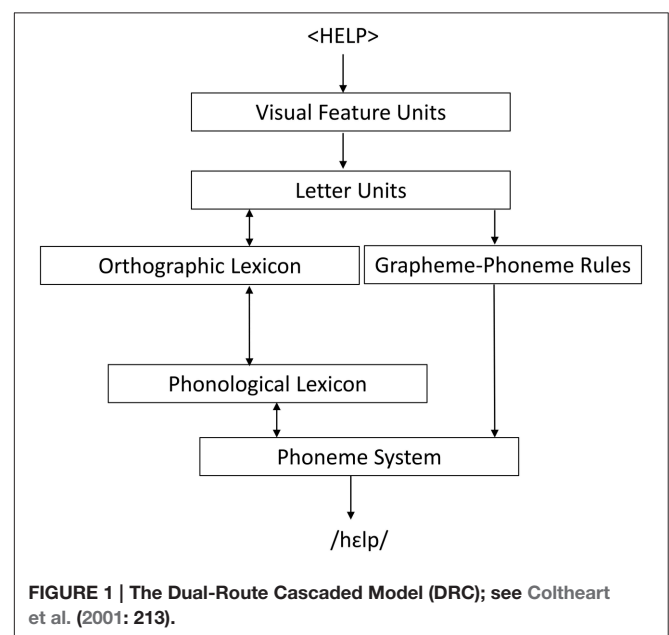
The mechanism of "affix stripping" which has been claimed to explain masked priming effects in English and other languages with alphabetic scripts only provides a partial account of the priming patterns for Japanese in our experiments. While the priming effect obtained for *-ta* suffixed past-tense forms in Experiment 1 is consistent with affix stripping, non-affixal /i/ forms were also found to yield a priming effect of a similar magnitude as *-ta* forms, even though /i/ is an epenthetic vowel rather than a morphological affix. We also observed a significant semantic priming effect in Experiment 1, again unexplainable in terms of affix stripping. Furthermore, Experiment 2 revealed a (kanji-mediated) priming effect for prime-target pairs which were morphologically completely unrelated, another finding that affix stripping cannot account for.

A second possibility we considered was that the decomposition mechanism that operates under masked-presentation conditions is *root-driven* in Japanese, unlike, for example, in English in which it is apparently affix-driven. That is, while English readers may try to identify potential affixes to be stripped off, Japanese readers may search for potential roots when

reading complex words. However, root spotting also provides only a partial account for our findings. While the similar-size priming effects for both the *-ta* and the /i/ conditions as well as the (smaller) semantic priming effect found in Experiment 1 can be attributed to overlapping or (semantically) related roots, the priming effect obtained in Experiment 2 cannot be explained in these terms, as all primes and targets tested here had unrelated roots.

This leaves us with the third possibility suggested above, that the priming effects obtained in the two experiments are orthographic in nature, due to the indirect orthographic overlap (in terms of kanji) between primes and targets. Recall, however, that all our stimuli were presented in moraic kana scripts. To see how kanji might be activated under such circumstances, consider our findings in the light of the dual-route reading model shown in **Figure 1**.

The DRC model assumes two alternative routes for the processing of written words, a lexical and a non-lexical one. If a stimulus is processed via the lexical route, the processor accesses the orthographic lexicon and activates a lexical entry for the word which corresponds to the orthographic properties of the particular stimulus. This entry is connected to a corresponding entry in the phonological lexicon which contains information about the phonological properties of the particular word. If a stimulus is instead processed via the non-lexical route, the processor does not access the orthographic lexicon, but instead directly activates a phonological representation of the stimulus on the basis of sign-sound correspondence rules. The DRC model has also been applied to Japanese. Coltheart (2014) proposed that in reading the common mixed script of Japanese, the lexical route is used for the logographic (kanji) components and the non-lexical route for the "syllabograms" of the two moraic scripts. A DRC account of Japanese reading and writing has also been used to explain why Japanese



patients with aphasia are often selectively impaired for kana, with performance for kanji remaining intact (e.g., Sasanuma and Fujimura, 1972; Sasanuma, 1975). As can be seen from **Figure 1**, while the DRC account posits two distinct pathways for reading, the lexical and the non-lexical routes, these two routes are indirectly linked through the phoneme system and the phonological lexicon. Subliminal kanji activation as found in our experiments can be explained through these links. Consider one of our prime words spelled with hiragana and suppose that this is read via the non-lexical reading route. The hiragana syllabograms then activate corresponding phonemes in the “phoneme system” and these, through the links shown in **Figure 1**, activate entries in the phonological lexicon and subsequently the orthographic lexicon. As entries in the Japanese orthographic lexicon also contain information about how a given phonological string is spelled with kanji, the corresponding kanji are also (albeit indirectly) activated. In this way, a kanji-mediated priming effect may arise even for stimuli entirely presented in moraic scripts.

Indirect kanji activation may not only account for why affixal and non-affixal prime words produced priming effects of a similar magnitude in our Experiment 1, but may also shed new light on findings from previous masked morphological priming studies on Japanese. Both Clahsen and Ikemoto (2012) and Fiorentino et al. (2015) found significant priming effects for word forms with both the productive (*-sa*) and the unproductive (*-mi*) nominalization suffix, which they interpreted in terms of affix stripping. Note, however, that the critical prime-target pairs used in both these studies had the same kanji overlap as the *-ta* and */i/* forms tested in the current Experiment 1. Indeed, in Fiorentino et al. (2015), primes and targets were even written in the mixed script (with kanji) when presented on screen. Thus, the priming effects reported in these studies are not necessarily morphological in nature, but may also be explainable through indirect (in the case of Clahsen and Ikemoto, 2012) or direct (in the case of Fiorentino et al., 2015) kanji activation.

Finally, we also found a significant semantic priming effect in Experiment 1, an unusual finding given previous studies which have typically not obtained semantic priming effects under masked priming conditions. It is conceivable that the semantic priming effect we found for Japanese is also a reflection of its particular writing system, in the following way. Recall that the mixed script (with kanji) is the common way of reading and writing in Japanese, and that kanji are logograms which represent words or roots and engage the lexical reading route; see also Perfetti et al. (2007) for evidence suggesting that logograms activate distinct brain reading networks. Coltheart (2014) notes that kanji are ‘indivisible wholes that are not composed of subword-level orthographic elements’. Arguably, the processor can directly retrieve semantic information from such logograms. As Japanese readers are used to reading through this lexical route, semantic information might be more directly and perhaps

more quickly accessible for them than readers of alphabetic scripts who are more used to reading via the non-lexical route. We acknowledge, however, that this final consideration remains speculative and that further research is needed to determine how different writing systems affect semantic effects under masked priming conditions.

In conclusion, the experimental results reported here should not be taken to mean that the Japanese language comprehension system does without morphological decomposition or without affix stripping. Instead, our results on Japanese, in comparison to those on English and languages with other alphabetical scripts, suggest that language-particular properties, in the present case differences between their writing systems, modulate the way morphologically complex words are processed during reading.

## AUTHOR CONTRIBUTIONS

YN/YI: Substantial contributions to the acquisition, analysis, or interpretation of data for the work; Drafting the work or revising it critically for important intellectual content; Final approval of the version to be published; Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. GJ/HC: Substantial contributions to the conception or design of the work; Drafting the work or revising it critically for important intellectual content; Final approval of the version to be published; Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## FUNDING

Alexander-von-Humboldt Professorship (HC), Scientific Grant-in-Aids (C) (No. 24520484, Yoko Nakano) and Kwansei Gakuin University Research Grant (A) (YN). We also acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam for covering Frontiers’ publishing charges.

## ACKNOWLEDGMENTS

We thank all participants and our students who helped us with the experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00316>

## REFERENCES

- Amano, N., and Kondo, K. (2003). *NTT Database Series: Lexical Properties of JAPANESE (Version 2) [CD-ROM]*. Tokyo: Sanseido.
- Boudelaa, S., and Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: incremental masked priming in Arabic. *Lang. Cogn. Process.* 20, 207–260. doi: 10.1080/01690960444000106
- Chen, H.-C., Yamaguchi, 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
- Clahsen, H., and Ikemoto, Y. (2012). The mental representation of derived words: an experimental study of *-sa* and *-mi* nominals in Japanese. *Ment. Lex. 7*, 147–182. doi: 10.1075/ml.7.2.02cla
- Clahsen, H., and Neubauer, K. (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua* 120, 2627–2637. doi: 10.1016/j.lingua.2010.06.007
- Coltheart, M. (2014). The neuronal recycling hypothesis for reading and the question of reading universals. *Mind Lang.* 29, 255–269. doi: 10.1111/mila.12049
- 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
- Davis, M., and Rastle, K. (2010). Form and meaning in early morphological processing: comment on Feldman, O'Connor, and Moscoso del Prado Martín (2009). *Psychon. Bull. Rev.* 17, 749–755. doi: 10.3758/PBR.17.5.749
- Dehaene, S. (2009). *Reading in the Brain: The New Science of How We Read*. New York, NY: Penguin.
- Duñabeitia, J., Perea, M., and Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition* 105, 691–703. doi: 10.1016/j.cognition.2006.12.001
- Feldman, L., Milin, P., Cho, K., Moscoso del Parado Martín, F., and O'Connor, P. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Front. Hum. Neurosci.* 9:111. doi: 10.3389/fnhum.2015.00111
- Feldman, L., O'Connor, P., and Moscoso del Parado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: a violation of form-then-meaning accounts of word recognition. *Psychon. Bull. Rev.* 16, 684–691. doi: 10.3758/PBR.16.4.684
- Fiorentino, R., Naito-Billen, Y., and Minai, U. (2015). Morphological decomposition in Japanese de-adjectival nominals: masked and overt priming evidence. *J. Psycholinguist. Res.* doi: 10.1007/s10936-015-9349-3. [Epub ahead of print].
- 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
- Heyer, V., and Clahsen, H. (2015). Late bilinguals see a scan in scanner AND in scandal: dissecting formal overlap from morphological priming in the processing of derived words. *Bilingualism* 18, 543–550. doi: 10.1017/S1366728914000662
- Kazanina, N., Dukova-Zheleva, G., Geber, D., Kharlamov, V., and Tonciulescu, K. (2008). Decomposition into multiple morphemes during lexical access: a masked priming study of Russian nouns. *Lang. Cogn. Process.* 23, 800–823. doi: 10.1080/01690960701799635
- Kiyose, G. N. (1971). Renketushiin to renketuboin to—Nihongo doshi mukatuyoron. [Juncture consonants and juncture vowels—a theory on no conjugation in Japanese]. *Kokugogaku* 86, 42–56.
- Longtin, C., and Meunier, F. (2005). Morphological decomposition in early visual word processing. *J. Mem. Lang.* 53, 26–41. doi: 10.1016/j.jml.2005.02.008
- Marslen-Wilson, W. D. (2007). “Morphological processes in language comprehension,” in *The Oxford Handbook of Psycholinguistics*, ed G. Gaskell (Oxford: Oxford University Press), 157–193.
- Neubauer, K., and Clahsen, H. (2009). Decomposition of inflected words in a second language: an experimental study of German participles. *Stud. Second Lang. Acquis.* 31, 403–435. doi: 10.1017/S0272263109090354
- Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., and Tan, L. H. (2007). Reading in two writing systems: accommodation and assimilation of the brain's reading network. *Bilingualism* 10, 131–146. doi: 10.1017/S1366728907002891
- Rastle, K., Davis, M., Marslen-Wilson, W., and Tyler, L. (2000). Morphological and semantic effects on visual word recognition: a time-course study. *Lang. Cogn. Process.* 15, 507–537. doi: 10.1080/0169096050119689
- Rastle, K., Davis, M., and New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychon. Bull. Rev.* 11, 1090–1098. doi: 10.3758/BF03196742
- Saito, H. (1981). Kanji to Kana no yomi ni okeru keitaitekifugoka oyobi onintekifugoka no kento. [Use of graphemic and phonemic encoding in reading Kanji and Kana.] *Jpn. J. Psychol.* 52, 266–273.
- Sampson, G. (1985). *Writing Systems: A Linguistic Introduction*. Stanford, CA: Stanford University Press.
- Sasanuma, S. (1975). Kana and Kanji processing in Japanese aphasics. *Brain Lang.* 2, 369–383. doi: 10.1016/S0093-934X(75)80077-0
- Sasanuma, S., and Fujimura, O. (1972). An analysis of writing errors in Japanese aphasic patients: kanji versus kana words. *Cortex* 8, 265–282. doi: 10.1016/S0010-9452(72)80034-0
- Shirota, S. (1998). *Nihongo Keitairon [Japanese Morphology]*. Tokyo: Hituji Shobo.
- Silva, R., and Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: evidence from masked priming experiments in English. *Bilingualism* 11, 245–260. doi: 10.1017/s1366728908003404
- Stanners, R., Neiser, J., Hernon, W., and Hall, R. (1979). Memory representation for morphologically related words. *J. Verbal Learn. Verbal Behav.* 18, 399–412. doi: 10.1016/S0022-5371(79)90219-6
- Tagawa, T. (2012). “Bunsan Keitairon o motiita dousikatuyou no kenkyuu ni mukete: Renyookei no bunpu ni okeru keitaitougoronteki mondai [For the study of verbs based on the theory of distributed morphology],” in *Katuyoron No Zensen [The Frontline of Theories of Conjugation]*, eds Y. Nitta and T. Masuoka (Tokyo: Kuroshio), 191–216.
- Thierry, G., and Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12530–12535. doi: 10.1073/pnas.0609927104

**Conflict of Interest Statement:** 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 © 2016 Nakano, Ikemoto, Jacob and Clahsen. 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) or licensor 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.



# Metaphoric Reference: An Eye Movement Analysis of Spanish–English and English–Spanish Bilingual Readers

Roberto R. Heredia\* and Anna B. Cieślicka

Cognitive Neuroscience Laboratory, Department of Psychology and Communication, Texas A&M International University, Laredo, TX, USA

This study examines the processing of metaphoric reference by bilingual speakers. English dominant, Spanish dominant, and balanced bilinguals read passages in English biasing either a figurative (e.g., describing a *weak and soft fighter that always lost and everyone hated*) or a literal (e.g., describing a *donut and bakery shop that made delicious pastries*) meaning of a critical metaphoric referential description (e.g., “creampuff”). We recorded the eye movements (first fixation, gaze duration, go-past duration, and total reading time) for the critical region, which was a metaphoric referential description in each passage. The results revealed that literal vs. figurative meaning activation was modulated by language dominance, where Spanish dominant bilinguals were more likely to access the literal meaning, and English dominant and balanced bilinguals had access to both the literal and figurative meanings of the metaphoric referential description. Overall, there was a general tendency for the literal interpretation to be more active, as revealed by shorter reading times for the metaphoric reference used literally, in comparison to when it was used figuratively. Results are interpreted in terms of the Graded Salience Hypothesis (Giora, 2002, 2003) and the Literal Salience Model (Cieślicka, 2006, 2015).

**Keywords:** anaphoric metaphor, bilingual metaphor, language dominance, metaphoric reference, referential metaphor

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Dawn Blasko,  
Penn State Erie, The Behrend  
College, USA  
Rachel Helen Messer,  
Oklahoma State University, USA

### \*Correspondence:

Roberto R. Heredia  
rheredia@tamiu.edu

### Specialty section:

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

**Received:** 01 August 2015

**Accepted:** 11 March 2016

**Published:** 29 March 2016

### Citation:

Heredia RR and Cieślicka AB (2016)  
Metaphoric Reference: An Eye  
Movement Analysis  
of Spanish–English  
and English–Spanish Bilingual  
Readers. *Front. Psychol.* 7:439.  
doi: 10.3389/fpsyg.2016.00439

## METAPHORIC REFERENCE: AN EYE MOVEMENT ANALYSIS OF BILINGUAL READERS

This study examines the comprehension of literal and non-literal meanings of a metaphoric description by Spanish–English and English–Spanish bilinguals at the discourse level. As a way to introduce the topic of this paper, consider the interpretation of a somewhat novel non-literal expression by a bilingual speaker as a professor remarked, “Back then in the late 1980s, I used to write poetry. Now I only write sleeping pills.” To which a bilingual student responded, “You mean prescription drugs?” The professor’s intended meaning (i.e., non-literal interpretation) was that he used to write fun and interesting stuff, and now he was writing books that were making people fall asleep. However, the student interpreted the expression literally as “sleeping pills,” which are typically related to drugs used to induce sleeping. “Sleeping pills,” in this case, is an example of a metaphoric referential description or an anaphoric metaphor (e.g., Gibbs, 1990; Onishi and Murphy, 1993; Budiu and Anderson, 2002; Stewart and Heredia, 2002; Almor et al., 2007; Heredia and Muñoz, 2015).



Anaphoric metaphors, or metaphoric referential descriptions, contrast with the conventionalized nominal metaphor of the form *A is B* (e.g., “books are sleeping pills”), where “books” is the *tenor* or *subject* (A) of the metaphor; “sleeping pills” is the *vehicle* that provides attributes characterizing the topic, and the *ground* is the basis on which it is possible to infer a relationship between the subject and the vehicle. Unlike in the nominal metaphor, the subject and vehicle in metaphoric referential descriptions occur apart from each other. The vehicle (i.e., “sleeping pills”) is used as the reference for the previously mentioned or implied subject (“books”). Thus, understanding referential descriptions requires that readers or listeners establish a connection between the anaphoric metaphor (“sleeping pills”) and its antecedent (“books”) found elsewhere in the sentence. This reactivation of the antecedent from the metaphoric reference, as argued by Gibbs (1990), Onishi and Murphy (1993), and Almor et al. (2007) requires additional inferential processes to understand the intended meaning (but see Heredia and Muñoz, 2015; see also Stewart and Heredia, 2002).

Although our knowledge about how bilinguals comprehend non-literal language in such domains as idiomatic expressions is on the rise (see for example, Heredia and Cieślicka, 2015), very few studies have looked at how bilingual speakers comprehend metaphoric referential descriptions in their second language (L2). The purpose of the present study is to look at early- vs. late-lexical processing of literal vs. non-literal language interpretation using eye movement recordings (Rayner, 2009), and to assess the effects of language dominance in Spanish-English and English-Spanish bilinguals (cf. Johnson and Rosano, 1993; Heredia, 1997; Cieślicka et al., 2014; cf. Cieślicka, 2015).

## Models of Metaphoric Processing

Two general theoretical models have been traditionally proposed in the monolingual literature to account for how metaphoric expressions are comprehended. These models have been extended to explain bilingual figurative language processing (e.g., Matlock and Heredia, 2002; Heredia and Muñoz, 2015; Vaid et al., 2015). The *Direct Access Model* assumes that during the course of comprehending a metaphoric expression, its intended or non-literal interpretation may be accessed directly “without first requiring an initial literal interpretation computed and rejected” (Blasko and Connine, 1993, p. 295; Glucksberg, 2001; see also Vaid et al., 2015). Although a literal interpretation may be temporarily available to construct the non-literal meaning of a metaphor, it is neither obligatory nor required before the metaphoric comprehension begins (Blasko and Connine, 1993, p. 295). For the *Indirect Processing Model* (Searle, 1979; Swinney and Osterhout, 1990), the metaphor’s literal interpretation is obligatory, and only if the literal interpretation is defective or does not fit the context, a search for a non-literal interpretation is initiated (Heredia and Muñoz, 2015). Although the empirical evidence supports the Direct Access Model, studies utilizing more sensitive online methods measuring language processing in real time have unequivocally supported the Indirect Processing Model (see for example, Janus and Bever, 1985; Swinney and Osterhout,

1990). However, other studies (e.g., Blasko and Connine, 1993; Heredia and Muñoz, 2015) utilizing similar methodologies, have reported findings difficult to reconcile by both theoretical frameworks.

A third model of metaphoric processing is the *Graded Saliency Hypothesis* [GSH; Giora (2003; see also, Cieślicka, 2006, 2015)]. Briefly, the GSH assumes that metaphoric expressions are best understood depending on which meaning (i.e., literal vs. figurative) is more salient. *Salient meanings* are readily accessible, excitable, and influenced by such factors as word frequency, familiarity, conventionality, and prototypicality/stereotypicality. *Non-salient meanings*, on the other hand, are less frequently used, less familiar, and would take longer to be triggered, requiring extra-inferential processes (Giora, 2003, p. 491). Thus, for an English native speaker, the figurative meaning of a highly familiar metaphor such as *loneliness is a desert* would be highly salient and readily interpretable as relating to a feeling of isolation. The literal meaning, related to “sand” or “dryness,” would be less salient, resulting in longer response times. The opposite would be true for low familiarity metaphoric expressions. In this case, the literal meaning of the metaphoric expression would be more salient (see for example, Blasko and Connine, 1993), thus allowing explaining the literal interpretation presented in the introductory example, where the student in question understood “sleeping pills” as a narcotic prescription medication.

Although the GSH has not been applied to bilingualism or L2 metaphoric processing, it could be argued that *meaning saliency* (i.e., a metaphor’s literal vs. non-literal interpretation) could also be influenced by L2/bilingual factors, such as language dominance (Heredia, 1997; Dunn and Fox Tree, 2009; Heredia and Cieślicka, 2014), language proficiency, linguistic environment (i.e., bilingual vs. monolingual community), first language (L1), and L2 age of acquisition, and acquisition context (i.e., whether L2 was acquired at home vs. school). Thus, for a bilingual speaker who is highly proficient and dominant in the L2, figurative interpretations of highly familiar metaphoric expressions, like for monolinguals, would be highly salient and readily accessible (e.g., Heredia and Muñoz, 2015). However, for less proficient learners of an L2, literal reading of a metaphoric expression would take precedence and become more salient (see for example, *Literal Salient Model*; Cieślicka, 2006, 2015).

## Bilingual Metaphoric Language Processing

How do bilinguals comprehend metaphoric expressions? Although limited, experimental evidence seems to suggest that, like monolinguals, bilingual speakers might be able to directly access the intended (figurative) meaning of a metaphoric expression, as predicted by the Direct Access Model. In one of the first studies, Nelson (1992) investigated memory for metaphor by non-fluent bilinguals. Spanish-English and French-English bilinguals translated metaphorical and literal expressions from Spanish (e.g., *Un árbol es un paraguas* or *un árbol es fuerte*) or French (*Un arbre est un parapluie* or *Un arbre est fort*) into

English (*A tree is an umbrella* or *A tree is strong*). For this condition, participants were explicitly instructed to translate the figurative or literal meaning of a metaphoric expression into English, the L2. In another experimental condition, bilinguals received a list of metaphors and literal expressions and were asked to simply translate them into L2. This condition did not require participants to consciously trigger the figurative interpretation of the metaphoric expression. A cued-recall task (e.g., *A tree is\_\_*) in Spanish or French was used to measure retention. Results revealed that translating the figurative meaning of a metaphor into L2 significantly improved retrieval, relative to translating the literal meaning of a metaphoric expression. More significant, however, was the finding that translating the figurative meaning of the metaphor did not produce better recall than the condition in which participants were simply asked to translate a metaphoric expression into L2. These results were interpreted as suggesting that the processing of the figurative meaning of a metaphor is *automatic* and that the processing of the literal interpretation of a metaphor is not obligatory (see Glucksberg and Keysar, 1990; see also Vaid et al., 2015). In fact, asking participants to interpret the literal meaning of a metaphor actually *interfered* with normal processing, thus resulting in poor recall performance (Nelson, 1992). By and large, studies interpreted as supporting the Direct Access Model involve nominal metaphors of the type *A is B*.

More recently, Heredia and Muñoz (2015) utilized the cross-modal naming task (Swinney, 1979; Blasko and Connine, 1993; Tabossi, 1996; Heredia and Blumentritt, 2002; Cieślicka, 2006; Cieślicka and Heredia, 2016) to explore the temporal course of meaning activation (i.e., literal vs. figurative) of metaphoric referential descriptions. Highly fluent bilinguals in English, the L2, from a predominantly English-Speaking area (Experiment 1) and Spanish-English/English-Spanish bilinguals from a bilingual community (Experiment 2) listened to story passages as the one described in (1) below.

- (1) Stu went to see the Saturday night fights. There was one fighter that Stu hated. This guy always lost. Just as the match was about to start, Stu went to get some snacks. He stood in line for 10 min. When he returned, the bout had been canceled. “What happened?” Stu asked a friend. His friend replied, “Aw, the creampuff<sub>[\*1]</sub> didn’t even show<sub>[\*2]</sub> up, I can’t believe it!”

As participants listened to the passage, a visual target appeared either at metaphor offset for Experiments 1 and 2 (position \*1 depicted by subscripts), 1000 ms (Experiment 1), or 300 ms (Experiment 2) after metaphor offset (position \*2). The visually presented target words were either related literally (e.g., “pastry”) or non-literally (e.g., “boxer”) to the metaphoric referential description (“creampuff”), or they were unrelated controls (“pirate” and “camel”).

At issue was whether “creampuff,” the vehicle of the metaphor (i.e., the anaphor), would (re)activate its antecedent (“fighter,” the non-literal meaning) and its literal interpretation (“pastry”). The priming effect was taken as a measurement of meaning (literal vs. non-literal) activation, or the extent to which a particular

meaning is activated relative to its control. Briefly, the priming effect refers to the robust finding whereby response to a target (e.g., “bread”) is faster when preceded by a related (e.g., “butter”) than an unrelated word (e.g., “mirror”).

Results from Experiment 1 showed that bilinguals living in a predominantly English environment were able to (re)activate the antecedent (“boxer”) immediately after the metaphoric referential description (Position 1). In contrast, there was no evidence of literal meaning activation. However, at 1000 ms, only literal activation was evident. Although bilinguals had direct access to the non-literal interpretation, as in Stewart and Heredia’s (2002) English monolingual speakers, activation of the literal interpretation suggested the literal interpretation remained as a possibility for bilinguals, even 1000 ms after they had accurately resolved the linguistic ambiguity (cf. Blasko and Connine, 1993). Experiment 2 indicated that, regardless of the target position, both the literal and non-literal meanings of the metaphoric referential description were equally accessible, particularly 300 ms after metaphor offset. However, it appeared that the literal interpretation was relatively more active. Heredia and Muñoz’s (2015) results were more consistent with the GSH suggesting that bilinguals, like native speakers of English, might have direct access to the metaphoric figurative interpretation. However, the overall evidence from these two experiments points to the possibility that literal meanings are more strongly coded or more salient in the bilingual’s lexicon (e.g., Cieślicka, 2006, 2015; cf. Vaid et al., 2015).

## The Present Study

How do bilinguals store and process metaphoric expressions? The purpose of the present study is to further investigate the on-line comprehension of referential metaphoric expressions using eye movement recordings, which provide temporally precise measures, to capture fine-grain differences, if any, between the literal and figurative interpretations of a metaphoric expression. If temporal differences (i.e., early vs. late processing stages) exist between literal and figurative meaning interpretations, as predicted by Direct and Indirect Processing Models, eye movement recordings reflecting early (e.g., first fixation duration or duration of the very first fixation on a word), and late stages (e.g., total reading time or the sum of all fixation durations) of lexical processing (e.g., Rayner, 1998, 2009; Raney and Bovee, 2015; Whitford et al., 2015) will reveal these processing differences. It may very well be the case that findings in which figurative meanings take precedence over literal interpretations (e.g., Stewart and Heredia, 2002; Heredia and Muñoz, 2015, Experiment 1) reflect late stages of lexical processing or semantic integration. Another aim of the present study is to investigate the effects of language dominance (i.e., fluency, lexical and syntactic knowledge, ease of accessibility; Dunn and Fox Tree, 2009) on literal and figurative meaning activation in metaphoric processing. Previous studies have only looked at bilingual directionality in relation to the bilingual’s L1 (Spanish or English) and L2 (English or Spanish). In fact, bilingual models of word recognition (e.g., Kroll and Stewart, 1994) hypothesize retrieval differences between L1-L2 and L2-L1. However, language dominance may be a better predictor of

lexical access, regardless of bilingual directionality (e.g., Heredia and Muñoz, 2015).

Participants in the present study read short passages such as (1) above, describing a weak boxer being referred to as a “creampuff” (the metaphoric condition). For the literal condition, unlike the original studies (e.g., Gibbs, 1990; Onishi and Murphy, 1993), the same target was utilized. However, the preceding context was biased toward the literal interpretation of the critical target (e.g., “creampuff”), as in (2) below:

- (2) Stu and his buddy went to the donut shop. There was a baker who made delicious pastries. Just before they ordered, Stu went to the bathroom. When he came back his buddy was outside the bakery shop. “What happened?” Stu asked, and his friend replied, “Aw, the creampuff wasn’t even that good! I can’t believe it.”

The present study differs from Gibbs (1990) and Onishi and Murphy (1993), in that participants are presented with full passages (as in 1 and 2 above), instead of seeing each passage line by line. Another notable difference is that, for both literal and metaphoric conditions, the same critical target (“creampuff”) is used, thus controlling for word level effects. At issue is whether the figurative interpretation of the metaphoric referential description (“creampuff” as a weak boxer in passage 1) is faster and more readily accessible during early stages of reading comprehension, as predicted by the Direct Access Model, or whether the literal sense (“creampuff” as a pastry in passage 2) has precedence over the figurative interpretation. In this case, eye movement recordings for early stages of reading comprehension (e.g., first fixation duration) will reveal faster reading times for the literal interpretation. However, the figurative meaning interpretation of the metaphoric reference should be faster during the late stages, as reflected by such eye movement measures as, for example, total reading time of lexical processing. In relation to language dominance, English dominant bilinguals are expected to have direct access to the figurative interpretation of the metaphoric referential description, as in Stewart and Heredia (2002) and Heredia and Muñoz (2015). However, Spanish dominant bilinguals are likely to conform to a language processing configuration in which the literal meaning is more salient during both early and late stages of reading comprehension, as predicted by the GSH (Giora, 2003), and Cieślicka’s (2006, 2015) Literal Salience Model.

To our knowledge, no eye-tracking studies have explored the effects of language dominance in the comprehension of metaphoric referential expressions or nominal metaphor by bilinguals. At the monolingual level, there is only one study using eye movements to explore processing differences between literal and non-literal interpretation in metaphoric processing. Inhoff et al. (1984) explored the effects of prior context (biased vs. non-biased) in the comprehension of nominal metaphors. Participants read passages that were either literally biased, (3) *in the back of the barn, the farmer’s youngest child gathered pebbles and skipped them deftly across a puddle by the chicken coop. He knew that he was supposed to be feeding the animals but*

*he kept on flicking at the birds. The hens clucked noisily;* literally non-biased (4) *In the back of the barn, the hens clucked noisily;* (5) non-literally biased, *at a meeting of the women’s club the youngest member requested the floor and brought up the issue of supporting the equal rights amendment. The importance of the issue outweighed her discomfort in speaking before the group. They reacted as she expected. The hens clucked noisily;* and non-literally non-biased, (6) *At a meeting of the women’s club, the hens clucked noisily.* Results from the Inhoff et al. (1984) study revealed no differences between literal and non-literal targets under biased contextual conditions. Under non-biased contextual conditions, literal targets were recognized more quickly than non-literal targets. It should be noted, however, that Inhoff et al.’s (1984) eye movement measures (e.g., total reading time and sentence reading time) reflected late stages of reading comprehension, and it is unclear if their results can be generalized to early processing stages of reading comprehension. To sum up, we ask the following questions: (1) what are the effects of language dominance in metaphoric processing? (2) Will there be differences between early and late stages of reading comprehension between literal and figurative interpretations of metaphoric referential expressions among bilinguals?

## MATERIALS AND METHODS

### Participants

Forty Spanish-English and 32 English-Spanish bilinguals (female = 55, male = 17) from the psychology subject pool at Texas A&M International University participated in the experiment. Participants volunteered or received class credit as a partial class requirement. Four participants were excluded from further analysis due to computer errors; one other participant was excluded because the language questionnaire was unavailable. Language dominance was assessed by Dunn and Fox Tree’s (2009) Bilingual Dominance Scale. Based on their aggregated scores, 20 participants were classified as “balanced” ( $M = -1.05$ ,  $SD = 2.87$ ), 36 as English dominant ( $M = 14.9$ ,  $SD = 4.80$ ), and 16 as Spanish dominant ( $M = -14.4$ ,  $SD = 6.32$ ). The participant’s mean age was 23.3 years ( $SD = 4.02$ ,  $range = 19-38$ ). Following Vaid et al. (2015), a composite language-proficiency score was created that included speaking, reading, understanding, and writing Spanish ( $M = 5.56$ ,  $SD = 1.25$ ) and English ( $M = 6.42$ ,  $SD = 0.9$ ). A dependent *t*-test revealed that participant’s proficiency scores were higher for English, [ $t(71) = 4.77$ ,  $p < 0.01$ ].

**Table 1** summarizes participants’ responses to language performance measures broken down into bilingual directionality (English-Spanish vs. Spanish-English) and English and Spanish. Both bilingual groups (English-Spanish vs. Spanish-English) reported similar formal education in Spanish and English. Likewise, both groups reported similar incidences of language mixing or code-switching (see for example, Heredia and Altarrriba, 2001), where both languages are used simultaneously. Mean self-ratings for English usage on a typical day and language fluency in English and Spanish for speaking,

**TABLE 1 | Language background information for the bilingual sample.**

	English–Spanish	Spanish–English
Age	24.1 (5.29)	22.5 (2.50)
Mean age L2 learned	5.81 (2.30)	6.68 (3.92)
Mean years of schooling in Spanish	4.03 (6.07)	6.48 (6.00)
Mean years of schooling in English	13.6 (5.54)	11.2 (5.68)
Mean language mixing Ratings	4.75 (2.03)	5.18 (1.87)
<b>Mean Self-Ratings</b>	<b>English</b>	<b>Spanish</b>
Language usage	6.11 (1.32)	5.25 (1.68)**
Speaking	6.23 (1.07)	5.68 (1.1)**
Reading	6.46 (0.948)	5.54 (1.41)**
Understanding	6.61 (0.707)	6.11 (1.33)**
Writing	6.35 (1.14)	4.86 (1.71)**

Values in parentheses represent SDs; \*\* $p < 0.01$ ; Self-ratings (1 = Not fluent, 7 = Very fluent); Language mixing (1 = I Never mix languages, 7 = I Mix languages all the time).

reading, understanding, and writing show that bilinguals rated English as the most frequently used and more proficient language.

**Table 2** summarizes correlations ( $r$ ) between variables typically used to measure language proficiency and language dominance. It is noteworthy that language usage and language proficiency are significantly correlated with language dominance. The overall pattern, depicted in **Table 2**, suggests that, as language use, language proficiency, and language dominance increase for one language, the other language decreases in the same indicators. Another important finding is that language mixing is positively correlated with Spanish and English proficiency. More notable, however, is the moderate correlation between proficiency and dominance for both languages and the strong negative relationship between language dominance in Spanish and English.

## Materials and Design

Stimuli consisted of 40 short passages, as described in 1–2 in the introduction, each a brief exchange between two persons describing a mutually known person or thing. Figurative-biased passages (see passage 1 above) were taken directly from Stewart

and Heredia (2002) and Heredia and Muñoz (2015). Literal-biased passages were created following the same format as the figurative-biased ones. For both passages, the target item (e.g., “creampuff”) appeared in the penultimate or last sentence, and it was preceded by a description of a very weak boxer (figurative condition) or a pastry or donut that was not that good after all (literal condition). For both conditions, sentences containing the critical target were constructed as similarly as possible (e.g., figurative: “Aw, the creampuff didn’t even show up, I can’t believe it” vs. literal: “Aw, the creampuff wasn’t even that good! I can’t believe it”). Average number of words for the metaphorical ( $M = 64.1, SD = 10.4$ ) and literal passages ( $M = 61.6, SD = 11.5$ ) was the same,  $t(78) = 1.03, p = 0.31$ ; likewise, the average number of words before the target location between literal ( $M = 50.2, SD = 10.7$ ) and figurative ( $M = 53.3, SD = 10.0$ ) passages was comparable,  $t(78) = 1.33, p = 0.19$ . Average number of words for the target location within each passage was 54.3 ( $SD = 9.95$ ), for the figurative and 51.2 ( $SD = 10.7$ ) for the literal conditions. The 3.1 word difference between the two passages was not statistically significant,  $t(78) = 1.33, p = 0.19$ .

Two lists were required to counterbalance each critical target within a passage. Each list contained 20 figurative- and 20 literal-biased passages. Lists were constructed in such a way that no passage with the same target (literal or non-literal) was repeated within a list. Stimuli assignment was between lists using an ABBA BAAB counterbalancing procedure. Forty passages were used as fillers. These passages were taken from Stewart and Heredia (2002) and Heredia and Muñoz (2015). Fillers were matched to the experimental stimuli on format and number of sentences and contained no metaphorical reference or any hint of figurative language (e.g., *Holy wanted to be the first female admitted to the male-only military academy. On her way to the academy, she noticed she didn’t have her teddy bear. “Stop!” she yelled to the driver. “You have to go back, I forgot my teddy bear and I need it to keep me company and to support me.” The driver responded very angrily, “Forget it, I am not going back!”*) The 80 passages were combined in a pseudo-random order, which imposed the constraint that no more than three experimental conditions occurred consecutively. Five additional filler passages served as

**TABLE 2 | Summary of intercorrelations for scores on language variables.**

Variables	AgeE	AgeS	S_sch	E_sch	Mix	S_use	E_Use	S_prof	E_prof	E_dom
AgeS	-0.09									
S_sch	0.30*	-0.15								
E_sch	-0.22	0.15	0.09							
Mix	0.01	-0.24*	-0.06	0.05						
S_use	0.35**	-0.38**	0.24*	-0.06	0.57**					
E_use	-0.36**	0.148	-0.42**	0.20	0.23	-0.12				
S_prof	0.38**	-0.35**	0.19	-0.18	0.27*	0.65**	-0.21			
E_prof	-0.33**	0.03	-0.34**	0.22	0.24*	-0.07	0.68**	-0.01		
E_dom	-0.70**	0.30*	-0.24*	0.32**	-0.03	-0.41**	0.57**	-0.42**	0.50**	
S_dom	0.60**	-0.47**	0.33**	-0.28*	0.04	0.41**	-0.54**	0.480**	-0.42**	-0.83**

AgeE/S, age learned English and Spanish; S\_sch/E\_sch, years of schooling in Spanish/English; Mix, code-switching; S\_use/E\_use, Spanish/English usage in a typical day; S\_prof/E\_prof, Spanish/English proficiency; E\_dom/S\_dom, English/Spanish dominance; \* $p < 0.05$ . \*\* $p < 0.01$ .

practice trials. After each passage, participants responded to a true/false comprehension question. There were a total of 80 questions.

## Design

The critical values measured were first fixation duration (the length of time the eyes spend on the target word the very first time they land on it), gaze duration (the sum of the duration of all fixations made on the word prior to exiting the word), total reading time (the sum of all fixation durations made on the target word, including re-reading), and go-past time (the sum of all fixation durations, which starts with the first fixation on the word up to the time the eyes fixate to the right of the word), both for the figurative (passage 1) and literal (passage 2) targets. The design conformed to a 2 (target type: figurative vs. literal)  $\times$  3 (language dominance: Spanish vs. English vs. balanced) mixed factorial design, with target type as a within-subjects factor, and language dominance as a between-subjects variable.

## Procedure

Upon arrival to the laboratory, participants completed a pencil and paper consent form. They were then instructed to sit comfortably, so that they were able to position their chin on a chinrest and maintain stability. Visual calibrations were conducted to ensure that the eye tracker was accurately recording the participants' eye movements. The visual recording was monocular, where the eye tracker recorded the right eye.

Experimental stimuli were presented using SR-Experiment Builder running on Windows OS 7, and the EyeLink 1000 eye tracker was connected to a dedicated host computer running on DOS. Participants were randomly assigned to one of the two experimental lists. They were seated approximately 55 cm from the monitor, with their head supported by a chinrest. Following eye-tracking calibration, the instructions were displayed on the computer screen. Participants were asked to read each passage shown on the screen and to answer a comprehension question that followed. Comprehension questions were of the yes/no type and were randomly displayed after every few sentences to ensure participants were attending to the passages. Participants were provided with 10 practice trials to familiarize them with the experimental procedure. Each trial started with a black fixation point appearing on the left of the screen, where the first word of the passage would appear. Participants were instructed to focus their eyes on the fixation point and to press the designated button on a Microsoft SideWinder Plug and Play Game Pad (Model GP5) game controller device in order to trigger the sentence display. After reading each passage, participants pressed the game controller button to advance to the next trial. Passages were displayed in black Times New Roman 20 font against a white background. The eye monitoring session lasted approximately 30 min. Following the experiment, participants completed the Bilingual Dominance Scale language background questionnaire and were debriefed as to the purpose of the experiment. The experimental protocol was approved by the Texas A&M International University Institutional Review Board (IRB).

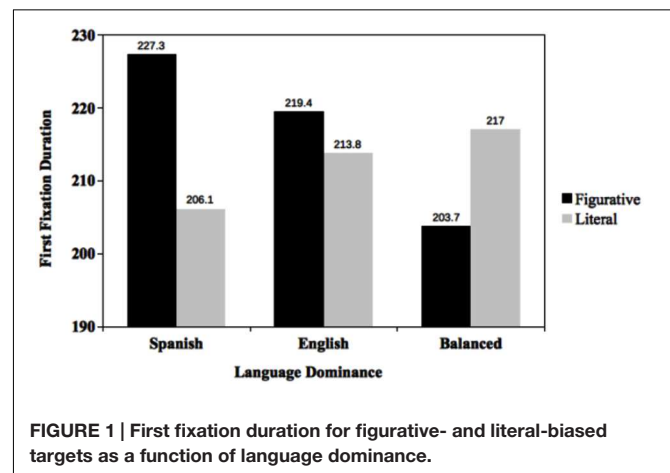
## RESULTS

Participants' responses to comprehension questions were analyzed for accuracy. All participants answered the comprehension questions with an accuracy above 90%, with the exception of one participant whose accuracy was 85%. Responses were normally distributed across the two experimental conditions. Data from four participants were excluded due to computer errors. Data from one additional participant were excluded due to an incomplete language questionnaire. The data were analyzed using linear mixed effects models (LME) using IBM SPSS V.20, mixed linear models procedure, with fixed (i.e., independent variables; target type and language dominance) and random effects (i.e., items and subjects). Analyses were conducted on both early (first fixation and gaze duration) and late (total reading time and go past duration) stage reading measures (Rayner, 1998). For all the measures, percentage of data removed and percentage of targets skipped as a function of the experimental conditions are provided.

### First Fixation Duration

A total of 3.4% of the data were removed because fixation durations were less than 100 ms (Libben and Titone, 2009). The LME 2  $\times$  3 analysis yielded a statistically reliable interaction between language dominance as a function of target type,  $F(2,1482.2) = 5.01, p < 0.01$ . **Figure 1** summarizes the interaction. Follow up *F-tests* show that Spanish dominant bilinguals read literal targets faster than figurative ones,  $F(2,595.0) = 6.0, p < 0.05$ . However, English dominant bilinguals were equally fast in reading both literal and figurative targets. Balanced bilinguals, on the other hand, were faster in reading figurative than literal targets; however, the reading differences were not significant,  $F(2,437.6) = 2.93, p = 0.09$ . No other effects reached significance.

The eye movement recordings showed a high proportion of participants skipping the target region. A higher rate of literal targets ( $M = 0.474, SE = 0.029$ ) was skipped in comparison to figurative targets ( $M = 0.380, SE = 0.029$ ),  $F(1,83.1) = 7.74, p < 0.01$ . Skipping a word during reading is reflective of the



word's predictability, or the participant's ability to anticipate the target word based on the preceding contextual constraints (e.g., Ehrlich and Rayner, 1981; Balota et al., 1985; Rayner and Well, 1996; Rayner et al., 2011). It is likely that literal targets, in this case, were easy to predict (and hence skipped) based on the preceding contextual information that was readily integrated, as opposed to the figurative target that required additional inferential processes (Giora, 2002; see also Onishi and Murphy, 1993; Almor et al., 2007).

## Gaze Duration

A total of 2.5% of the data were removed because gaze durations were less than 100 ms. The two-way analysis revealed a significant main effect of target type,  $F(1,82.1) = 3.84$ ,  $p = 0.05$ . No other effects were reliable. However, it should be noted that, although the interaction was not statistically reliable, gaze duration patterns are similar to the first fixation data, showing the larger differences between literal and figurative targets for Spanish dominant bilinguals. In relation to target type, literal targets ( $M = 261.9$ ,  $SE = 12.03$ ) exhibited shorter reading times than figurative targets ( $M = 289.3$ ,  $SE = 11.8$ ).

In relation to skipping the target region as a function of language dominance and target type, the only reliable main effect was target type,  $F(1,82.9) = 7.56$ ,  $p < 0.01$ , where the proportion of skipped literal targets ( $M = 0.470$ ,  $SE = 0.029$ ) was higher than figurative targets ( $M = 0.377$ ,  $SE = 0.029$ ).

## Go-Past Duration

A total of 1.7% of the data were removed because fixation durations were less than 100 ms. The two-way analysis revealed a main effect of target type,  $F(1,82.01) = 4.54$ ,  $p < 0.05$ . Literal targets ( $M = 299.1$ ,  $SE = 15.3$ ) were read faster than figurative targets, ( $M = 335.7$ ,  $SE = 15.0$ ). In regards to skipping the critical target, the two-way interaction revealed a main effect on target type,  $F(1,82.8) = 8.04$ ,  $p < 0.01$ . No other effects reached significance. The proportion of skipping a literal target ( $M = 0.469$ ,  $SE = 0.029$ ) was larger than for figurative targets ( $M = 0.373$ ,  $SE = 0.029$ ).

## Total Reading Time

A total of 1.8% of the data were removed because fixation durations were less than 100 ms. The two-way analysis revealed a main effect of target type,  $F(1,80.257) = 7.96$ ,  $p < 0.01$ . No other effects reached statistical significance. Literal targets ( $M = 395.8$ ,  $SE = 25.1$ ) were read faster than figurative targets ( $M = 477.0$ ,  $SE = 25.0$ ). Likewise, the two-way analysis on skipping the target during reading showed a main effect of target type,  $F(1,81.5) = 4.71$ ,  $p < 0.05$ . No other effects were significant. Literal targets ( $M = 0.190$ ,  $SE = 0.023$ ) exhibited larger proportions of skipping than figurative targets ( $M = 0.133$ ,  $SE = 0.023$ ).

## DISCUSSION

English dominant, Spanish dominant, and balanced bilinguals read passages biasing either a figurative (e.g., describing a

*weak and soft fighter that always lost and everyone hated*) or a literal meaning (e.g., describing a *donut and bakery shop that made delicious pastries*) of a metaphoric referential description. For both conditions, we recorded the eye movements for the critical anaphoric reference (e.g., "creampuff" in the boxing/pastry scenarios). We utilized first fixation and gaze durations known for their sensitivity to tap into early stages of lexical processing, and go-past duration and total reading time measurements which reflect late-stage, post-lexical processing. At issue was whether bilinguals would access the figurative interpretation of the metaphoric referential description during the "automatic" or early stages (i.e., first fixation, gaze duration) of reading comprehension, as hypothesized by the Direct Access Model. Alternatively, the literal interpretation should have precedence in the early stages, whereas figurative meaning should emerge in late stages, as posed by the Indirect Processing Model.

The results revealed that, at least for first fixation durations, meaning activation (figurative vs. literal) was moderated by language dominance (see **Figure 1**). Spanish dominant bilinguals yielded shorter reading times for literal than figurative interpretations, suggesting that, based on the overwhelming use of Spanish, the literal representation of the metaphoric expression was more readily accessible. English dominant and balanced bilinguals, on the other hand, have access to both interpretations. This suggests that at early stages of metaphoric processing, English dominant and balanced bilinguals are considering both the literal and figurative interpretations. In relation to existing models of figurative language processing, at least during early stages of bilingual figurative language comprehension, the results are inconsistent with models that predict direct access to the intended (i.e., figurative) meaning only. Neither are the current data consistent with the indirect models predicting literal meaning activation first. Our results indicate that meaning activation is moderated by language dominance and meaning salience (Giora, 2002), in which the literal meaning seems to be more salient than figurative meaning (e.g., Cieślicka, 2006, 2015), especially for Spanish dominant bilinguals.

Gaze duration measurements failed to replicate the findings from first fixation duration. There was a tendency for the literal interpretation to be more active, as revealed by shorter reading times as compared to the figurative interpretation. This finding was generalized throughout both the go-past duration and total reading time measures, as well as skipping rates, in which literal targets were more likely to be skipped than figurative ones. To summarize, gaze duration, go-past duration and total reading time (i.e., measures hypothesized to reflect late stage, post-lexical processing such as semantic integration, revision, problem solving) replicate the original findings reported by Gibbs (1990) and others (Onishi and Murphy, 1993; Budiu and Anderson, 2002; Almor et al., 2007; but see Stewart and Heredia, 2002; Heredia and Muñoz, 2015), showing that the literal interpretations of metaphoric referential descriptions are read faster, as they are easier to process than figurative interpretations. These finding generalize to both monolingual and bilingual experiments investigating

reading processes (cf. Inhoff et al., 1984, Experiment 1). These results point to a possible bilingual model of metaphoric processing that is moderated by meaning salience, in which *literal salience* takes precedence over figurative salience (Cieślicka, 2006, 2015).

Taken together, the evidence from the present study qualifies Heredia and Muñoz (2015) results showing that multiple meaning activation occurs, but only for English dominant and balanced bilinguals who are active in English, and only during the early stages of lexical processing. As originally reported by Gibbs (1990), Onishi and Murphy (1993), Budiu and Anderson (2002), and Almor et al. (2007), the present results also show that the literal anaphoric reference to the antecedent (i.e., “creampuff” referring to a pastry) is read faster than metaphoric anaphoric reference (i.e., “creampuff” referring to a coward boxer), but only at late processing stages, as measured by go-past duration and total reading time (cf. Inhoff et al., 1984).

The present results, showing retention of literal meaning at late processing stages, can be addressed within the framework of the *retention hypothesis*, which supplements the GSH (Giora, 2002) and which explains the activation of contextually incompatible meanings in the course of metaphorical comprehension. According to the retention hypothesis, contextually incompatible meanings accessed initially on account of their salience may, subsequently, be either maintained or suppressed, depending on their contribution to the utterance interpretation. Accordingly, inappropriate meanings which are conducive to the compatible interpretation are retained, whereas meanings conflicting with the compatible meaning are discarded. Since literal meanings may sometimes contribute to the construction of figurative interpretations, in the figurative-biasing context, the incompatible literal meaning may remain active even after the contextually appropriate figurative interpretation has been determined, as long as it is in some way supportive of the figurative interpretation. On the other hand, in the literally biasing context, where the contextually appropriate meaning is the literal one, the incompatible figurative meaning is usually irrelevant to the construction of the literal utterance meaning and hence becomes quickly suppressed.

Suppression of the contextually incompatible meaning postulated by Giora (2002) in the retention hypothesis is congruent with the suppression mechanism posed to play a crucial role in the understanding of figurative language by Gernsbacher and Robertson (1999), who define suppression as *a general cognitive mechanism, the purpose of which is to attenuate the interference caused by the activation of extraneous, unnecessary, or inappropriate information* (Gernsbacher and Robertson, 1999, p. 1619). Suppression has been experimentally demonstrated to attenuate inference during lexical access of ambiguous words, the processing of anaphoric and cataphoric reference, syntactic parsing, as well as in the understanding of metaphorical language (e.g., Gernsbacher and Robertson, 1999; Gernsbacher et al., 2001; Glucksberg et al., 2001).

Predictions of the retention hypothesis were tested by Giora and Fein (1999) in a word fragment completion test, which measured the amount of activation of literal and

figurative meanings in literally and figuratively biasing contexts. Participants were first presented with short stories, ending with the target figurative sentence and then completed fragmented words which were related to either the literal or the figurative meaning of the target sentence. For each target sentence, two short texts were constructed, one biasing the literal meaning of the figurative target and the other biasing its metaphorical interpretation. For example, the literal biasing context for the target *Only now did they wake up* was a story about people partying and dancing all night, and some people calling on their friends the day after the party with the friends opening the door half asleep. In turn, the metaphorically biasing context for the same target sentence was a story about a bloody war going on in central Europe in which thousands of innocent lives had been lost before a decision was made to intervene and put an end to the massacres. In accordance with the authors' predictions, familiar metaphorical expressions activated both their literal and figurative meanings in both types of context, with the figurative meaning retained to a significantly smaller extent in the literally than in the figuratively biasing context. According to Giora and Fein (1999), these results supported the view that in the literally biasing context with which it is incompatible, the non-literal meaning of a figurative expression gets suppressed very quickly, whereas in the figuratively biasing context the literal meaning of a figurative expression is retained because of its relevance for utterance processing. Overall, results reported in the study described here are compatible with the retention hypothesis, as both literal and figurative meanings were found active in dominant and balanced bilinguals, and literal meanings were retained even in contextually incompatible figurative-biased utterances.

What can existing models say about bilingual metaphoric processing? Although previous findings (e.g., Nelson, 1992; Vaid et al., 2015) have found evidence that bilinguals, like monolinguals, have direct access to the metaphor's figurative interpretation, as hypothesized by Direct Access Models, the present results are more consistent with Giora's (2003) GSH and Cieślicka's (2006, 2015) Literal Salience Model. As revealed by the two-way interaction for the first fixation duration reading measure, language dominance moderates which meaning (literal vs. figurative) is more salient, and as argued by Cieślicka (2006, 2015), there is a propensity for the literal meaning of metaphoric expressions to be more readily accessible and more salient for bilingual speakers.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

This research was supported in part by grant P031M105048 from the U.S. Department of Education to RH, and a National Science Foundation (NSF) Major Research Instrumentation (MRI) Grant

(BCS-1229123) to AC. The authors would like to thank Victoria I. Peña and Patricia González for their invaluable assistance in the development of the stimuli and data collection. As always, we are

grateful to Omar García for his comments on the first draft, and to two anonymous reviewers whose comments and suggestions greatly improved the quality of the paper.

## REFERENCES

- Almor, A., Arunachalam, S., and Strickland, B. (2007). When the creampuff beat the boxer: working memory, cost, and function in reading metaphoric reference. *Metaphor Symbol* 22, 169–193. doi: 10.1080/10926480701235478
- Balota, D., Pollatsek, A., and Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cogn. Psychol.* 17, 364–390. doi: 10.1016/0010-0285(85)90013-1
- Blasko, D. G., and Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *J. Exp. Psychol.* 19, 295–308.
- Budiu, R., and Anderson, J. R. (2002). Comprehending anaphoric metaphors. *Mem. Cogn.* 30, 158–165. doi: 10.3758/BF03195275
- Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by L2 speakers. *Second Lang. Res.* 22, 115–144. doi: 10.1191/0267658306sr2630a
- Cieślicka, A. B. (2015). “Idiom acquisition and processing by second/foreign language learners,” in *Bilingual Figurative Language Processing*, eds R. R. Heredia and A. B. Cieślicka (New York, NY: Cambridge Press), 208–244.
- Cieślicka, A. B., and Heredia, R. R. (2016). “Priming and online multiple language activation,” in *The Bilingual Mind and Brain Series*, Vol. 1, *Methods in Bilingual Reading Comprehension Research*, eds R. R. Heredia and A. B. Cieślicka (New York: Springer), 123–156.
- Cieślicka, A. B., Heredia, R. R., and Olivares, M. (2014). “The eyes have it: how language dominance, salience, and context affect eye movements during idiomatic language processing,” in *Essential Topics in Applied Linguistics and Multilingualism. Studies in Honor of David Singleton*, eds L. Aronin and M. Pawlak (Berlin: Springer), 21–42.
- Dunn, A. L., and Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. *Bilingualism* 12, 273–289. doi: 10.1017/S1366728909990113
- Ehrlich, S. F., and Rayner, K. (1981). Contextual effects on word recognition and eye movements during reading. *J. Verb. Learn. Verb. Behav.* 20, 641–655. doi: 10.1016/S0022-5371(81)90220-6
- Gernsbacher, A. M., Keysar, B., Robertson, R. R. W., and Werner, N. K. (2001). The role of suppression and enhancement in understanding metaphors. *J. Mem. Lang.* 45, 433–450. doi: 10.1006/jmla.2000.2782
- Gernsbacher, A. M., and Robertson, R. R. (1999). The role of suppression in figurative language comprehension. *J. Pragmat.* 31, 1619–1630. doi: 10.1016/S0378-2166(99)00007-7
- Gibbs, R. W. (1990). Comprehending figurative referential descriptions. *J. Exp. Psychol.* 16, 56–66.
- Giora, R. (2002). Literal vs. figurative language: different or equal? *J. Pragmat.* 34, 487–506. doi: 10.1016/S0378-2166(01)00045-5
- Giora, R. (2003). *On Our Mind: Salience, Context, and Figurative Language*. New York, NY: Oxford University Press.
- Giora, R., and Fein, O. (1999). On understanding familiar and less-familiar figurative language. *J. Pragmat.* 31, 1601–1618. doi: 10.1016/S0378-2166(99)00006-5
- Glucksberg, S. (2001). *Understanding Figurative Language*. New York, NY: Oxford Press.
- Glucksberg, S., and Keysar, B. (1990). Understanding metaphorical comparisons: beyond Similarity. *Psychol. Rev.* 97, 3–18. doi: 10.1037/0033-295X.97.1.3
- Glucksberg, S., Newsome, M. R., and Goldvarg, Y. (2001). Inhibition of the literal: filtering metaphor-irrelevant information during metaphor comprehension. *Metaphor Symbol* 16, 277–293. doi: 10.1207/S15327868M1603
- Heredia, R. R. (1997). Bilingual memory and hierarchical models: a case for language dominance. *Curr. Dir. Psychol. Sci.* 6, 34–39. doi: 10.1111/1467-8721.ep11512617
- Heredia, R. R., and Altarriba, J. (2001). Bilingual language mixing: why do bilinguals code-switch? *Curr. Dir. Psychol. Sci.* 10, 164–168. doi: 10.1111/1467-8721.00140
- Heredia, R. R., and Blumentritt, T. L. (2002). On-line processing of social stereotypes during spoken language comprehension. *Exp. Psychol.* 49, 208–221. doi: 10.1026//1618-3169.49.3.208
- Heredia, R. R., and Cieślicka, A. B. (2014). “Bilingual memory storage: compound-coordinate and derivatives,” in *Foundations of Bilingual Memory*, eds R. R. Heredia and J. Altarriba (New York, NY: Springer), 11–39.
- Heredia, R. R., and Cieślicka, A. B. (Eds). (2015). *Bilingual Figurative Language Processing*. New York, NY: Cambridge Press.
- Heredia, R. R., and Muñoz, M. E. (2015). “Metaphoric reference: a real-time analysis,” in *Bilingual Figurative Language Processing*, eds R. R. Heredia and A. B. Cieślicka (New York, NY: Cambridge Press), 89–116.
- Inhoff, A. W., Lima, S. D., and Carroll, P. J. (1984). Contextual effects on metaphor comprehension in reading. *Mem. Cogn.* 12, 558–567. doi: 10.3758/BF03213344
- Janus, R. A., and Bever, T. G. (1985). Processing of metaphoric language: an investigation of the three-stage model of metaphor comprehension. *J. Psychol. Res.* 14, 473–487. doi: 10.1007/BF01666722
- Johnson, J., and Rosano, T. (1993). Relation of cognitive style to metaphor interpretation and second language proficiency. *Appl. Psycholinguist.* 14, 159–175. doi: 10.1017/S014271640000953X
- 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: 10.1016/j.jecp.2008.10.004
- Libben, M. R., and Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *J. Exp. Psychol.* 35, 381–390. doi: 10.1037/a0014875
- Matlock, T., and Heredia, R. R. (2002). “Understanding phrasal verbs in monolinguals and bilinguals,” in *Bilingual Sentence Processing*, eds R. R. Heredia and J. Altarriba (North Holland: Elsevier), 251–274.
- Nelson, E. M. (1992). Memory for metaphor by nonfluent bilinguals. *J. Psycholinguist. Res.* 21, 111–125. doi: 10.1007/BF01067990
- Onishi, K. A., and Murphy, G. O. (1993). Metaphoric reference: when metaphors are not understood as easily as literal expressions. *Mem. Cogn.* 21, 763–772. doi: 10.3758/BF03202744
- Raney, G. E., and Bovee, J. C. (2015). “Reading integration in bilingual speakers,” in *The Bilingual Mind and Brain Book Series*, Vol. 1, eds R. R. Heredia, J. Altarriba, and A. B. Cieślicka (New York, NY: Springer).
- 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
- Rayner, K., Slattery, T. J., Drieghe, D., and Liversedge, S. P. (2011). Eye movements and word skipping during reading: effects of word length and predictability. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 514–528. doi: 10.1037/a0020990
- Rayner, K., and Well, A. (1996). Effects of contextual constraint on eye movements in reading: a further examination. *Psychon. Bull. Rev.* 3, 504–509. doi: 10.3758/BF03214555
- Searle, J. (1979). “Metaphor,” in *Metaphor and Thought*, ed. A. Ortony (New York, NY: Cambridge Press.), 92–123.
- Stewart, M. T., and Heredia, R. R. (2002). Comprehending spoken metaphoric reference: a real-time analysis. *Exp. Psychol.* 49, 34–44. doi: 10.1027/1618-3169.49.1.34
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *J. Verbal Learn. Verbal Behav.* 18, 645–659. doi: 10.1016/S0022-5371(79)90355-4
- Swinney, D. A., and Osterhout, L. (1990). Inference generation during auditory language comprehension. *Psychol. Learn. Motiv.* 25, 17–33. doi: 10.1016/S0079-7421(08)60246-7



- Tabossi, P. (1996). Cross-modal semantic priming. *Lang. Cogn. Process.* 11, 569–576. doi: 10.1080/016909696386953
- Vaid, J., López, B. G., and Martínez, F. E. (2015). “Linking the figurative to the creative: bilingual’s comprehension of metaphors, jokes, and remote associates,” in *Bilingual Figurative Language Processing*, eds R. R. Heredia and A. B. Cieślicka (New York, NY: Cambridge Press), 53–86.
- Whitford, V., Pivneva, I., and Titone, D. (2015). “Eye movement methods to investigate bilingual reading,” in *The Bilingual Mind and Brain Book Series: Vol 1. Bilingual Reading Comprehension Research*, eds R. R. Heredia, J. Altarriba, and A. B. Cieślicka (New York: Springer).

**Conflict of Interest Statement:** 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 © 2016 Heredia and Cieślicka. 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) or licensor 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.*



# Effects of Valence and Origin of Emotions in Word Processing Evidenced by Event Related Potential Correlates in a Lexical Decision Task

Kamil K. Imbir<sup>1\*</sup>, Tomasz Spustek<sup>2</sup> and Jarosław Żygierewicz<sup>2</sup>

<sup>1</sup> Faculty of Psychology, University of Warsaw, Warsaw, Poland, <sup>2</sup> Faculty of Physics, University of Warsaw, Warsaw, Poland

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Dwight A. Krehbiel,  
Bethel College, USA  
DeMond M. Grant,  
Oklahoma State University, USA

### \*Correspondence:

Kamil K. Imbir  
kamil.imbir@gmail.com

### Specialty section:

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

**Received:** 24 September 2015

**Accepted:** 12 February 2016

**Published:** 02 March 2016

### Citation:

Imbir KK, Spustek T  
and Żygierewicz J (2016) Effects  
of Valence and Origin of Emotions  
in Word Processing Evidenced by  
Event Related Potential Correlates in a  
Lexical Decision Task.  
*Front. Psychol.* 7:271.  
doi: 10.3389/fpsyg.2016.00271

This paper presents behavioral and event-related potential (ERP) correlates of emotional word processing during a lexical decision task (LDT). We showed that valence and origin (two distinct affective properties of stimuli) help to account for the ERP correlates of LDT. The origin of emotion is a factor derived from the emotion duality model. This model distinguishes between the automatic and controlled elicitation of emotional states. The subjects' task was to discriminate words from pseudo-words. The stimulus words were carefully selected to differ with respect to valence and origin whilst being matched with respect to arousal, concreteness, length and frequency in natural language. Pseudo-words were matched to words with respect to length. The subjects were 32 individuals aged from 19 to 26 years who were invited to participate in an EEG study of lexical decision making. They evaluated a list of words and pseudo-words. We found that valence modulated the amplitude of the FN400 component (290–375 ms) at centro-frontal (Fz, Cz) region, whereas origin modulated the amplitude of the component in the LPC latency range (375–670 ms). The results indicate that the origin of stimuli should be taken into consideration while deliberating on the processing of emotional words.

**Keywords:** valence, origin of emotion, duality of mind, word processing, lexical decision task

## INTRODUCTION

This paper contributes to research on event-related potential (ERP) correlates of emotional word processing. There are several dimensions of emotional quality of stimuli charged with affect (Osgood et al., 1957). The valence (Kissler et al., 2007, 2009; Herbert et al., 2008; Schacht and Sommer, 2009a,b), arousal (e.g., Cuthbert et al., 2000; Schupp et al., 2000, 2004; Dillon et al., 2006; Fischler and Bradley, 2006) and concreteness (e.g., Kanske and Kotz, 2007; Barber et al., 2013; Palazova et al., 2013) of words modulate ERP correlates of word processing. Recently, we showed that the origin of an affective state also influences emotional word processing (Imbir et al., 2015a). In this study we used a standard lexical decision task (LDT) to compare the processing of lexically meaningful words and formally similar (readable, multi-syllabic) pseudo-word stimuli (Imbir et al., 2015b). We were interested in two aspects of processing. Firstly, we decided to compare word and pseudo-word reading conditions to find differences attributable to lexical processing (e.g., Münte et al., 1997; Bentin et al., 1999). Secondly, we investigated differences in the brain correlates of processing of emotional words caused by involuntary - implicit - semantic processing. Based on

the emotion duality model (c.f. Jarymowicz and Imbir, 2015), we hypothesized that emotional word processing would be influenced by the origin of an emotional state included in word meaning.

## The Emotion Duality Model

The diversity of human emotions is as great as the number of people in the world (Kagan, 2007). Science searches for ways to organize this diversity and tries to explain what constitutes an emotion, rather than simply describing their diversity. For example the constructionist point of view (e.g., Russell, 2003) claims that the diversity of emotions derives from the human mind's creation of subjective emotional states based on *core affect* – a description of the state of an organism in terms of pleasantness (valence) and activation (arousal) – with engagement of specific mechanisms (e.g., attributions, constructions of objects and situations evoking emotions and so on). Each state contributes to subjectively perceived experience.

*Duality of mind* theories (e.g., Gawronski and Creighton, 2013) offer a different perspective on emotional diversity, contrasting automated and controlled processes. The recently proposed emotion duality model (Jarymowicz and Imbir, 2015) is based on the concept of duality of mind. It postulates the existence of two separate evaluative mechanisms, namely the automatic evaluating system (AES), based on direct evaluations of external stimulation, and the reflective evaluating system (RES), based on cognitive appraisals (c.f. Imbir et al., 2015a).

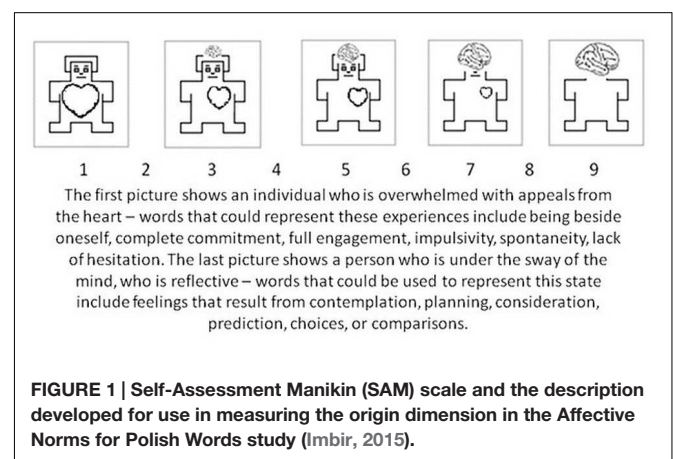
The concept of AES processing is based on the biological value criterion proposed by Damasio (2010), which stipulates that all organisms are driven to preserve their lives. Biological value criterion does not require language to appear; evaluations just happen, and they immediately influence the organism's state of mind, motivation and behavior. For example, sweetened water has a universal biological appeal because sugar is a source of energy, which is required to maintain life. Some stimuli have positive or negative biological significance (e.g., tasty food, appealing sexual partner, warm and sunny weather, snakes, nasty smells or decomposing corpses and so on). Our reaction to such stimuli may be immediate, because they have occurred repeatedly in human evolutionary past. Furthermore, these situations have shaped reproduction probability (making it less or more likely), thus organisms have evolved to react to biologically significant stimuli by approaching or avoiding them (Damasio, 2010).

Reflective evaluating system processing is based on verbalization (Jarymowicz and Imbir, 2015). Language is a crucial tool that allows our minds to organize stimuli effectively and gives us a temporal perspective (Rolls, 2000; Damasio, 2010) which includes both an imagined past and future. Language also increases our ability to distinguish emotional states, and to some extent allows us to modify our default biological responses. For example, fatty food may be evaluated as tasty and pleasant but, an individual who is trying to lose weight may adjust his or her evaluation to reflect this goal. It is possible that if one's motivation to lose weight is strong enough fatty food would be judged repulsive, rather than attractive. Use of linguistic evaluation criteria (Reykowski, 1989) relies on reflective processing, which is further based on propositional thinking

(Strack and Deutsch, 2004, 2014). This type of thinking requires effortful processing, but gives us an opportunity to modify automatic behavior in order to achieve a goal. Verbalization and use of language is such a frequent human activity that we can easily forget about or neglect its importance (c.f. Rolls, 2000). The constructionist theories of emotion (Russell, 2003), the appraisal theories (e.g., Scherer, 2005) and the emotion duality approach (e.g., Jarymowicz and Imbir, 2015) argue that language and controlled processes are crucial for understanding the diversity of emotions.

The origin of an emotional state (AES or RES) may modulate its influence on behavior. The basic assumption underlying this study is that origin, as a property of systems for processing emotional experience, is represented in language. This is supported by claims of the lexical hypothesis formulated in a field of personality psychology (e.g., Crowne, 2007). The lexical hypothesis states that characteristics important for people's lives will eventually become a part of their language, and should be represented in words and lexical structures describing them or associated with them. We think that the crucial distinction between automatic and controlled processing is likely to be represented in language (Imbir, 2015) and that even states which can arise without an associated linguistic representation will have a verbal label (e.g., pain). This was the rationale for our studies concerning affective norms for Polish words (Imbir, 2015, submitted). To capture AES and RES processing we used a Self Assessment Manikin (SAM) scale with a heart and mind metaphor (Imbir, 2015). This metaphor compares and contrasts (1) purely emotional processing “from the heart,” which is immediate and automatic, with (2) careful consideration and reflection “from the mind.” In order to clarify the SAM scale the additional descriptions of its meaning were presented at the top of the scale. This gave us more confidence that the participants would use the scale as intended and it resulted in very good reliability of estimations (c.f. Imbir, 2015). **Figure 1** presents the SAM scale and the descriptions provided to participants to clarify its meaning.

It is worth comparing the proposed origin dimension to the established dimension of concreteness since both, to some extent, describe the complexity of stimuli. Concrete stimuli are stimuli in



the physical world that we can touch or see, especially stimuli that can be easily represented by a picture or a word. Abstract stimuli are states that do not exist in the physical world (e.g., ideas or processes). These cannot be achieved without language or some sort of symbolic representation. It is especially hard to show them in one unambiguous picture. Automatic responses are immediate responses to stimulation. As in the case of a concrete stimuli, our mind labels AES emotional states, but they are able to appear even without language. Reflective responses are the product of RES-based appraisal and are based on language; they cannot occur without language. This makes concreteness and origin similar, but to verify this we analyzed affective ratings of 4900 Polish words of every grammatical class (Imbir, under review). In that study, participants assessed, among other variables, the origin of the state evoked by reading a given word, and the concreteness of the words. Origin and concreteness were only weakly correlated across the entire sample of words ( $r = -0.299$ ,  $p < 0.001$ ); they had only 10% common variance, which is low given the similarities mentioned above. On this basis we claim that origin and concreteness are distinct constructs, and that it is worth investigating effects of origin, controlled for variability in concreteness.

We think also that the origin of an emotional state is an important modulator of cognitive processes. This expectation is based on duality of mind theories (e.g., Gawronski and Creighton, 2013) which assert that many processes including decision making (e.g., Epstein, 2003; Kahneman, 2003; Darlow and Sloman, 2010), attitude formation (Gawronski and Creighton, 2013) and choices (Kahneman, 2011) are influenced by two types of 'mind.' The emotion duality model is an attempt to describe the outcomes of emotional processes for cognitive processes. The origin of stimuli was found to be related to ability to maintain cognitive control (Imbir and Jarymowicz, 2013), to modulate attention (Imbir, 2013) and to influence ERP correlates of word processing (Imbir et al., 2015a). Subjective significance, supposed to be the type of activation characteristic for reflective processes, was shown to eliminate the increase in response latencies caused by arousal in a modified Stroop task (Imbir, 2016). Taken together these results prompted us to investigate whether words with different origins would be processed differently in way that would be reflected in electroencephalography (EEG) correlates.

## Stages of Visual Processing of Emotional Words

The processing of visual words comprises several different stages (Bentin et al., 1999): the visual encoding of letters, translation of the letter shapes into a sequence of graphemes and orthographic patterns and, finally, activation of lexical and phonological structures and their meanings. Existing models of word processing (e.g., Rastle, 2007) do not include an affect as a salient factor in word recognition. Recently some evidence for an interaction between emotional valence and concreteness of a stimulus in word processing has emerged (e.g., Palazova et al., 2013). Effects of visual word processing on the N200, N400, FN400, and P600 (or LPC) ERP components have been found.

The N200 component evident in posterior locations is thought to be a manifestation of orthographic processing (e.g., Nobre et al., 1994) which distinguishes meaningful words, pronounceable pseudo-words and unpronounceable non-words from other complex, non-orthographic stimuli.

In anterior locations the N400 component is considered to be a manifestation of an unexpected event in speech perception and reading, such as when the last word in a sentence is not consistent with the sentence structure or meaning (e.g., Kutas et al., 1984). This N400 component is elicited only by words or pseudo-words that follow phonological rules (Bentin et al., 1999) and has been found even in paradigms which use isolated printed or spoken words rather than sentences as stimuli (e.g., Bentin et al., 1993). In frontal and central locations the N400 component represents higher order processing of semantic meaning and the word recognition stage of processing (c.f. Kutas and Federmeier, 2011) and it has also been shown to be sensitive to multiple lexical variables (Barber and Kutas, 2007) such as orthographic, phonological and word class information.

The FN400 component is similar to N400, but located in frontal regions (Kutas and Federmeier, 2011). Curran (2000) investigated the effects of familiarity and recollection of words. He showed that familiarity influenced the ERP waveform in the 300-500 ms interval in frontal location (FN400) whilst recollection was related to LPC (400-800 ms) in posterior locations. Further studies have demonstrated that FN400 is not specific to word processing (Voss and Paller, 2007); it also occurs in response to abstract geometric shapes which are being rated for perceived meaningfulness. In Voss and Paller (2007) study stimulus familiarity was manipulated as number of repetitions. Only meaningful shapes elicited an FN400 component, suggesting that this component is connected with conceptual priming rather than familiarity (Voss and Paller, 2007; Kutas and Federmeier, 2011).

The P600 component found in posterior locations is sometimes termed (e.g., Palazova et al., 2013) a late positive complex (LPC), and is considered to be a manifestation of deeper processing of word meaning (for a review see: Citron, 2012). The P600 component was found to be sensitive to processing difficulty and stimulus valence. Citron et al. (2011) reported that neutral words that were less salient than comparison valenced words elicited a larger LPC response during a LDT. Other higher order processes, such as perception of stimulus relevance (Fischler and Bradley, 2006), behavioral performance in the task (Polich, 2007), processing of self-referential (e.g., "my happiness") versus other-referential (e.g., "his/her success") words (Herbert et al., 2010, 2011) and processing of words differing in origin of emotional state (Imbir et al., 2015a) also modulated LPC amplitude.

It is worth emphasizing that emotional words are processed differently from other, more salient emotional stimuli. For example, although we might expect that attending to emotional stimuli would influence early ERP components - as processing faces or emotional scenes does - whereas in fact emotional word processing appears to modulate later ERP components associated with semantic analysis (Palazova et al., 2013). To conclude, existing evidence suggests that the emotional content

of a word modulates the N200 amplitude (in studies of emotional word processing the N200 component is often referred to as the early posterior negativity, EPN), as well as in the N400 and P600 (or LPC) components (for a review see Citron, 2012).

## The Origin of the Emotional State as a Factor Modulating the Processing of Words

Our previous ERP study of word processing (Imbir et al., 2015a) have shown that the amplitude of early (EPN) and late (LPC) ERP components is modulated by both the valence and origin of words. In that study participants performed an odd-ball task. They were instructed to decide whether a word presented to them had negative or positive connotations, and to avoid responding to a predefined standard word. Independent component analysis (ICA) revealed a specific independent component with dipolar fronto-occipital topography that corresponded to the modulation of EPN. It had a higher absolute amplitude for positive words than for negative words. LPC amplitude was more positive for stimuli of reflective rather than automatic origin. ICA also revealed a dipolar independent component located in the left parietal region (direction left-parietal – right-frontal) which had a higher absolute amplitude (in the 437-570 ms window) in the case of emotional stimuli which engaged the automatic system rather than the reflective system. This difference was not observed in processing of neutral words. The emotion-related difference in the left parietal independent component described above (Imbir et al., 2015a) was not specific to concreteness differences.

## Aims and Hypotheses

The aim of this study was to investigate how the valence and origin of stimuli influenced implicit, involuntary lexical processing of verbal stimuli in a LDT. Previous studies of word processing in a task which explicitly demanded lexical processing (Imbir et al., 2015a) convinced us that this would extend understanding of emotional word processing mechanisms. Our argument is that origin is an important property of words and that it is processed independently of valence.

We expected words to elicit stronger ERP responses than pseudo-words in time ranges considered sensitive to word processing (N200, FN400 and LPC components). Considering words conditions only we did not expect effects of valence and origin on N200 component as this component is thought to reflect early processing, and hence orthographic rather than semantic processing (e.g., Nobre et al., 1994). This is in contrast to tasks involving explicit processing of words, where EPN effects related to allocation of visual attention have been reported (c.f. Citron, 2012). We expected to find differences in involuntary semantic processing (not related to instructions or to the task) with respect to the FN400 and LPC components. We expected that valence-related differences would occur sooner (in the FN400 time range) whilst origin-related differences would occur later (in the LPC time range). We also expected that origin effects would be asymmetrical (c.f. Imbir et al.,

2015a) and evident mainly in the posterior regions of the left hemisphere.

## MATERIALS AND METHODS

### Participants

Thirty-two individuals (women = 15; men = 17) aged from 19 to 26 years ( $M = 21.5$ ,  $SD = 1.63$ ) participated in the study. They were students at various Warsaw colleges and universities and participated voluntarily for a small reward. All of the participants were right-handed, native Polish speakers with normal or corrected-to-normal vision. Participants provided verbal, informed consent to participation; we did not collect written consent as we had assured the participants of anonymity. Oral consent was provided in the presence of at least two members of the laboratory and documented by them in a research diary. This procedure was suggested by the bioethical committee which approved the research. We did not collect any personal data from our participants. The design, experimental conditions and consent procedure for this study were approved by the bioethical committee of the Maria Grzegorzewska University.

### Design

The study consisted of two stages: first we searched for differences between the processing of words and pseudo-words. At this stage, we applied two-factor repeated measures analysis of variance (stimulus type  $\times$  location) to successive time intervals. We assumed that activity in intervals where such differences are detectable is relevant to specific aspects of word processing. During the second stage responses to words were analyzed further by means of a three-factorial design: origin (automatic, A; no particular origin or mixed origin, 0; reflective, R)  $\times$  valence (negative, Neg; neutral, Neu; positive, Pos)  $\times$  scalp location. Combinations of conditions are referred to using the concatenation of their abbreviations, e.g., Neg\_0 represents a negatively valenced word of no particular origin. We controlled for variability in arousal, concreteness, length and frequency of words and for participants' handedness, gender and use of medication.

### Linguistic Material and its Properties Emotionally Charged Words

The linguistic material consisted of a set of nouns divided into nine groups of 15 words; groups were matched for arousal, concreteness, length, and frequency, but differed with respect to valence and origin, yielding a 3 (valence)  $\times$  3 (origin) factorial manipulation. The selection of words was based on a previous study (Imbir, under review) in which the valence, origin, arousal and concreteness of 4905 Polish words were assessed by at least 50 participants (25 women), studying at various Warsaw universities. The database study used the same methodology as a previous study to define the properties of 1586 Polish words (Imbir, 2015). For the different levels of valence and origin we selected words as follows, level 1: score at least 1  $SD$  below mean (Neg or A); level 2: score within 0.5  $SD$  of the mean (Neu

or 0); level 3: score at least 1 *SD* above the mean (Pos or R). All selected words had scores within 0.5 *SD* of the mean for arousal and concreteness. The selection procedure also ensured that the groups were matched for frequency and word length (NoL). Word frequency estimates were based on occurrence in a database of online Polish texts (Kazojć, 2011) and represent the number of times each word appeared in the database. The distribution of frequencies in this database was positively skewed so word frequency data were natural logarithm (LN) transformed to permit use of parametric statistics.

To check that our manipulations of word properties operated as intended we conducted 3 (valence) × 3 (origin) ANOVAs for each measured dimension. We found the predicted group differences in valence and origin ratings and an absence of group differences in arousal, concreteness and frequency. Neu and 0 words appeared to be about one letter shorter than the words in other groups. The full results of these analyses can be found in Appendix 1 (Word properties). **Table 1** presents means and standard deviations for word properties for all groups. See Appendix 2 for a complete list of selected words and their properties.

### Pseudo-Words

Pseudo-words were taken from the Polish Pseudo-words List (Imbir et al., 2015b), a dataset consisting of a large number of randomly generated pseudo-words stimuli assessed by competent

judges as fulfilling the criteria for pseudo-word stimuli. The 135 pseudo-words used in this study were selected from a list of 864 pseudo-word stimuli that were positively verified by all of five judges and matched the 135 word stimuli as closely as possible with respect to length (number of letters). Appendix 1 (Table A1) presents the list of pseudo-words used in the experiment.

### Procedure

Participants were informed about the aim of the experiment and nature of the EEG measurement. We encouraged them to maintain a comfortable posture and control their eye blinks. The protocol provided 3-s breaks for normal blinking every 10 trials, as well as two longer breaks, whose duration controlled by the participant, for rest and adjustment of posture. The long breaks occurred every 270 trials.

The task was to read stimuli as they appeared in the middle of the screen and to classify them as words or pseudo-words by pressing tagged keys on the keyboard. The content and latency of responses were recorded. A single experimental block comprised 135 words and 135 pseudo-words; this block was repeated three times. Words and pseudo-words were displayed in random order in all blocks. Trials proceeded as follows: (1) fixation point displayed for 500 ms; (2) stimulus displayed until participant responds; (3) blank screen displayed for randomly varied interval between 1000 and 1100 ms.

**TABLE 1 | Descriptive statistics (*M*, *SD*) for groups of words used in factorial manipulation.**

	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>	Origin category	<i>M</i>	<i>(SD)</i>
Valence	3.50	(0.36)	5.02	(0.56)	6.71	(0.35)	Automatic	5.07	(1.39)
Origin	4.45	(0.53)	4.58	(0.37)	4.33	(0.70)		4.45	(0.55)
arousal	4.37	(0.49)	4.15	(0.55)	4.28	(0.80)		4.27	(0.62)
concreteness	4.31	(1.15)	3.95	(0.74)	4.48	(1.20)		4.24	(1.05)
NoL	7.20	(2.65)	7.47	(1.96)	7.40	(2.41)		7.36	(2.31)
Ln_freq	5.21	(1.91)	5.65	(2.03)	5.73	(2.28)		5.53	(2.04)
Valence	3.37	(0.36)	5.19	(0.54)	6.38	(0.32)	Control (0)	4.98	(1.32)
Origin	5.41	(0.31)	5.49	(0.30)	5.36	(0.35)		5.42	(0.32)
arousal	4.15	(0.23)	4.12	(0.67)	4.04	(0.51)		4.11	(0.49)
concreteness	4.05	(1.12)	3.96	(1.32)	4.17	(0.74)		4.06	(1.06)
NoL	6.47	(2.03)	5.27	(1.33)	6.93	(2.02)		6.22	(1.92)
Ln_freq	5.48	(2.28)	5.97	(1.27)	6.61	(2.02)		6.02	(1.92)
Valence	3.66	(0.35)	5.30	(0.39)	6.49	(0.40)	Reflective	5.15	(1.23)
Origin	6.46	(0.30)	6.63	(0.41)	6.63	(0.56)		6.57	(0.43)
arousal	4.32	(0.49)	3.93	(0.47)	4.03	(0.36)		4.10	(0.46)
concreteness	4.17	(1.13)	4.09	(1.17)	4.41	(1.07)		4.22	(1.11)
NoL	7.07	(1.75)	6.27	(1.62)	7.20	(2.27)		6.84	(1.91)
Ln_freq	5.42	(1.37)	6.53	(1.79)	6.01	(1.22)		5.99	(1.52)
<b>Valence category</b>	<b>Negative</b>		<b>Neutral</b>		<b>Positive</b>			<b>Total</b>	
Valence	3.51	(0.37)	5.17	(0.50)	6.53	(0.38)	Total	5.07	(1.31)
Origin	5.44	(0.92)	5.57	(0.92)	5.44	(1.09)		5.48	(0.97)
arousal	4.28	(0.42)	4.07	(0.56)	4.12	(0.58)		4.16	(0.53)
concreteness	4.18	(1.11)	4.00	(1.08)	4.35	(1.01)		4.18	(1.07)
NoL	6.91	(2.15)	6.33	(1.86)	7.18	(2.20)		6.81	(2.09)
Ln_freq	5.37	(1.85)	6.05	(1.72)	6.12	(1.89)		5.85	(1.84)

## EEG Materials

### Apparatus

Stimuli were displayed on a standard PC monitor (LCD display; 15-inch diagonal). A second PC was used for monitoring and recording EEG data. Stimuli and EEG data were synchronized using a custom-made hardware trigger. EEG activity was recorded from 19 electrode sites, Fz, Cz, Pz, Fp1/2, F7/8, F3/4, T3/4, C3/4, T5/6, P3/4, O1/2, referenced to linked earlobes, grounded on the clavicle and with impedances of 5 k $\Omega$  or less. The signal was acquired using a Porti7 (TMSI) amplifier with a sampling frequency of 256 Hz.

### Offline EEG Signal Processing

Offline analysis was performed in Matlab<sup>®</sup> with the EEGLAB toolbox (Delorme and Makeig, 2004). The statistical tests were implemented using the appropriate R procedures (R Development Core Team, 2008, available from: <http://www.R-project.org>). The signal was zero-phase filtered with Butterworth low- and high-pass filters (second order: corresponding to 12 dB/octave roll-off, with half amplitude cut-off frequencies of 30 and 0.1 Hz respectively), and with an IIR notch filter to remove line noise at 50 Hz. To suppress activity common to most of the data channels, the data were re-referenced to common-average montage. Epochs from 300 ms pre-stimulus to 1000 ms post-stimulus were extracted and baseline-corrected (baseline data taken from -200 to 0 ms).

The data were visually inspected to exclude error and artifact trials (e.g., eye blinks or muscle activity) the mean number of artifact-free trials eliciting a correct classification response was as follows: word trials  $M = 338$  ( $SEM = 7$ ); pseudo-word trials  $M = 348$  ( $SEM = 7$ ). Paired sample  $t$ -tests revealed a difference between trial types,  $t(31) = 4.17$ ,  $p < 0.0003$ . For further analysis of ERP data in the words  $\times$  pseudo-words design, randomly selected trials of the more numerous stimulus type were removed on a per subject basis to achieve a sample in which there were equal numbers of error-free trials of each type (word; pseudo-word) for each subject, a procedure suggested by Thomas et al. (2004). After this equalization procedure the mean number of trials of each type was 335 ( $SEM = 7$ ).

The mean number of trials in per word condition was 37.5 ( $SEM = 0.3$ ). The Friedman test for replicated block design did not indicate differences in the average number of trials per condition for the valence groups with origin as a blocking variable [ $\chi^2(2) = 3.9$ ,  $p = 0.1$ ], or for the origin groups with valence as a blocking variable [ $\chi^2(2) = 4.7$ ,  $p = 0.1$ ].

## RESULTS

### Behavioral Measures

The distributions of behavioral measures, i.e., response time and response accuracy were non-normal according to a Shapiro-Wilk test for all word and pseudo-word categories. Differences between levels of experimental factors were therefore assessed with non-parametric tests.

### Response Times

Response times were longer for pseudo-words ( $M = 934$  ms,  $SEM = 44$  ms) than for words ( $M = 776$  ms,  $SEM = 28$  ms). The Wilcoxon signed-rank test yielded  $V = 523$ ,  $p < 5e-09$ . There were no effects of experimental factors on response latency for word trials. The Friedman test for replicated block design did not indicate differences in response times for valence groups with origin as a blocking variable [ $\chi^2(2) = 2.7$ ,  $p = 0.2$ ], or for origin groups with valence as a blocking variable [ $\chi^2(2) = 0.4$ ,  $p = 0.8$ ]. The mean response time was 777 ms ( $SEM = 10$  ms).

### Response Accuracy

There were more errors on pseudo-word trials [ $M = 4.1\%$  ( $SEM = 0.3\%$ )] than word trials [ $M = 3.1\%$  ( $SEM = 0.4\%$ )]; the Wilcoxon signed-rank test yielded  $V = 390$ ,  $p < 0.02$ . **Table 2** shows mean response accuracy and standard error for word trials in the valence  $\times$  origin design. The Friedman test for replicated block design revealed an effect of valence group with origin as a blocking variable [ $\chi^2(2) = 38.8$ ,  $p < 1e-08$ ], and an effect of origin group with valence as a blocking variable [ $\chi^2(2) = 14.8$ ,  $p < 0.001$ ].

Using the Wilcoxon pairwise test with the Holm correction for multiple comparisons we demonstrated that Pos words were more likely to be classified correctly than Neg ( $p < 2e-8$ ) and Neu words ( $p < 2e-6$ ). Zero words (control condition) were less likely to be classified correctly than A words ( $p = 0.05$ ) and R words ( $p < 0.003$ ). A *post hoc* test using the Wilcoxon rank sum test with the Holm correction showed that the highest number of errors was associated with Neg\_0 words, more errors were with Neg\_0 words than in all other conditions except Neu\_A words.

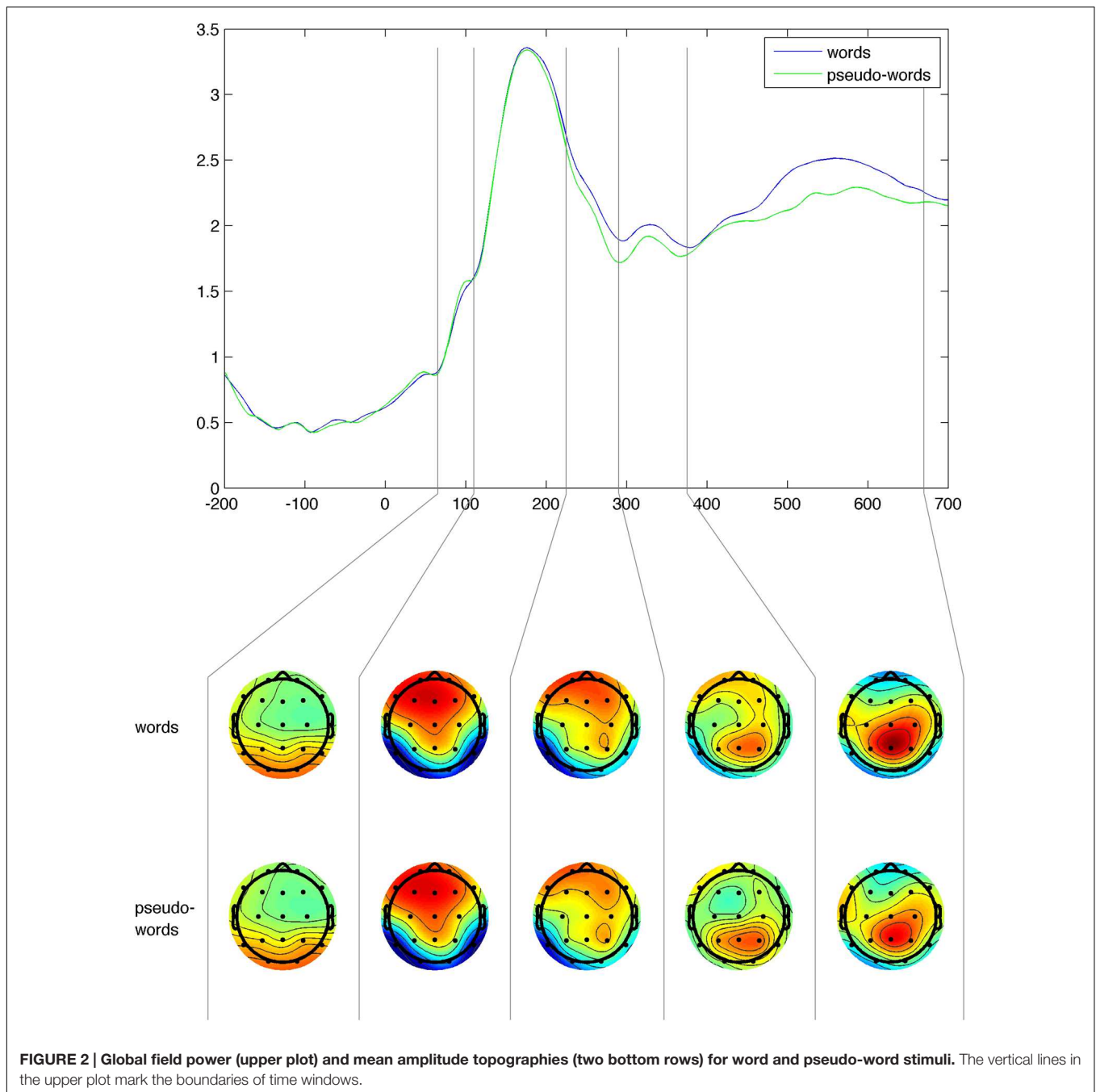
## Electrophysiological Data

### Time Windows and ROIs Selection

Event-related potential data were analyzed for the following time windows: 65–110, 110–225, 225–290, 290–375, and 375–670 ms, based on the global field power (GFP) curve (**Figure 2**). GFP is computed as spatial standard deviation, and quantifies the sum of electrical activity over all electrodes at a given time point. The latencies of GFP maxima indicate the latencies of evoked potential components (Lehmann and Skrandies, 1980; Skrandies, 1990). **Figure 2** shows that the amplitude topographies for the first two time-windows are very similar for words and pseudo-words. In the remaining three time windows there are differences between words and pseudo-words with respect to amplitude and distribution. In the fourth time window word stimuli produced larger amplitude responses in the frontal regions than pseudo-words. The time windows used in the analysis correspond also

**TABLE 2 | Percentage correct responses in ( $M$  and  $SEM$ ) for each stimulus category.**

	Neg	Neu	Pos	Total
A	96.8 (0.6)	96.0 (0.7)	98.4 (0.4)	97.1 (0.4)
O	91.0 (1.3)	96.0 (0.7)	99.1 (0.3)	95.4 (0.6)
R	98.0 (0.4)	97.8 (0.4)	99.0 (0.3)	98.3 (0.2)
Total	95.3 (0.6)	96.6 (0.4)	98.8 (0.2)	96.9 (0.3)



to these that were assigned a potential role in word processing (Bentin et al., 1999; Kutas and Federmeier, 2011).

We selected five regions of interest (ROIs): left frontal (Fp1 and F7), centro-frontal (Fz and Cz), right frontal (Fp2 and F8), left parietal (C3 and P3) and right parietal (C4 and P4). Signal amplitudes in these regions were averaged across the corresponding electrode sites. Those regions are specific to components of interest (FN400, LPC: c.f. Voss and Paller, 2007). Based on previous findings (Imbir et al., 2015a) we expected that origin effects would be lateralized and therefore investigated left, right and central ROIs. We used an ROI approach rather

than analyzing individual components at specific sites and time-windows suggested by the literature in order to not to bias the data analysis by subjective choices. The precise location of certain components varies between studies (e.g., Kutas and Federmeier, 2011) for methodological reasons. Our approach was based on the assumption that averaging activity from different sites in one ROI would allow us to identify the maximal for a given area response without subjectively choosing for analysis a single electrode. Also use of this approach was motivated by an incorporation of origin dimension, not examined earlier, thus potentially resulting in amplitude changes in different sites. On



the other hand consideration of all individual electrodes would augment the problem of multiple comparison.

### Analysis of Differences between Words and Pseudo-Words

We carried out separate repeated measures ANOVAs (stimulus type  $\times$  ROI) on mean amplitude for each time window. Amplitude was measured as the mean amplitude (averaged over the duration of the time window) as this is more robust against electrical noise and latency jitter than maximum amplitude in a given time window (Luck, 2005). There was a main effect of ROI in all time windows, but since only interactions with the location factor are meaningful in the case of average referenced data this finding is not discussed further. For the time windows where there was an effect of stimulus type *post hoc* analysis using Holm-corrected paired sample *t*-tests was used to identify the regions in which the effect was significant. Interactions between ROI and stimulus type are detailed below for each time window. No stimulus type effects were observed for time windows 65-110 ms and 110-225 ms.

#### 225-290 ms time window

There was a simple effect of stimulus type [ $F(1,31) = 17.84$ ,  $p < 0.0002$ ] and an interaction between stimulus type and ROI [ $F(4,124) = 2.89$ ,  $p < 0.03$ ]. *Post hoc* analysis revealed that response amplitude was more positive for words ( $M = 0.84 \mu\text{V}$ ,  $SEM = 0.21 \mu\text{V}$ ) than for pseudo-words  $M = 0.53 \mu\text{V}$  ( $SEM = 0.21 \mu\text{V}$ ) at the centro-frontal ROI [ $t(31) = 4.59$ ,  $p < 0.0004$ ].

#### 290-375 ms time window (corresponding with N400 or FN400)

There was a simple effect of stimulus type [ $F(1,31) = 64.09$ ,  $p < 5e-9$ ] and an interaction between stimulus type and ROI [ $F(4,124) = 14.82$ ,  $p < 7e-10$ ]. *Post hoc* analysis showed that all frontal ROIs words elicited more positive amplitudes than pseudo-words, but at the left-parietal ROI this pattern was reversed. The details are presented in Appendix 1 (Table A2).

#### 375-670 ms time window (corresponding with LPC)

There was a simple effect of stimulus type [ $F(1,31) = 8.17$ ,  $p < 0.008$ ] and an interaction between stimulus type and ROI [ $F(4,124) = 9.89$ ,  $p < 6e-7$ ]. *Post hoc* analysis showed that at the centro-frontal, left parietal and right parietal ROIs words elicited more positive amplitudes than pseudo-words, but at left frontal and right frontal ROIs the opposite pattern was observed. The details are presented in Appendix 1 (Table A3).

### Word Properties

We carried out separate three-factor repeated measures ANOVAs (valence  $\times$  origin  $\times$  ROI) on mean amplitude data for each of the three time windows in which there were differences between responses on word and pseudo-word trials. For significant effects two-way analysis of variance was performed at each ROI followed by *post hoc* analysis using Holm-corrected paired samples *t*-tests. Only reliable effects in each time window and ROI are reported below. The results are illustrated as ERP time courses in **Figures 3 and 4**.

**Figure 3** shows that in the 290-375 ms time window there was a clear positive deflection in the centro-frontal ROI in response to stimuli with Pos valence but not Neu or Neg valence. The same patterns is observable in the topographical distribution of contrast potentials.

**Figure 4** illustrates an interesting origin-related difference in the time course of the response in right-frontal and left-parietal ROIs in the 375-670 ms time window. The topographical distribution of amplitude contrast reveals that the differences between A and R stimuli and 0 stimuli follow a dipolar pattern. These observations are corroborated by the statistical analysis reported below.

#### 225-290 time window

No reliable effects were observed.

#### 290-375 time window (corresponding with N400 or FN400)

There was a three-way interaction between valence, origin and ROI [ $F(16,496) = 2.77$ ,  $p < 0.0003$ ]. There was a main effect of valence at the centro-frontal ROI [ $F(2,62) = 6.44$ ,  $p < 0.003$ ]. *Post hoc* analysis revealed that this was due to higher amplitude responses to Pos stimuli [ $M = 0.31 \mu\text{V}$ ,  $SEM = 0.14 \mu\text{V}$ ] than to Neu stimuli [ $M = -0.05 \mu\text{V}$ ,  $SEM = 0.15 \mu\text{V}$ ;  $t(95) = 3.2$ ,  $p < 0.004$ ] and Neg stimuli [ $M = -0.02 \mu\text{V}$ ,  $-SEM = 0.14 \mu\text{V}$ ;  $t(95) = 3.44$ ,  $p < 0.003$ ].

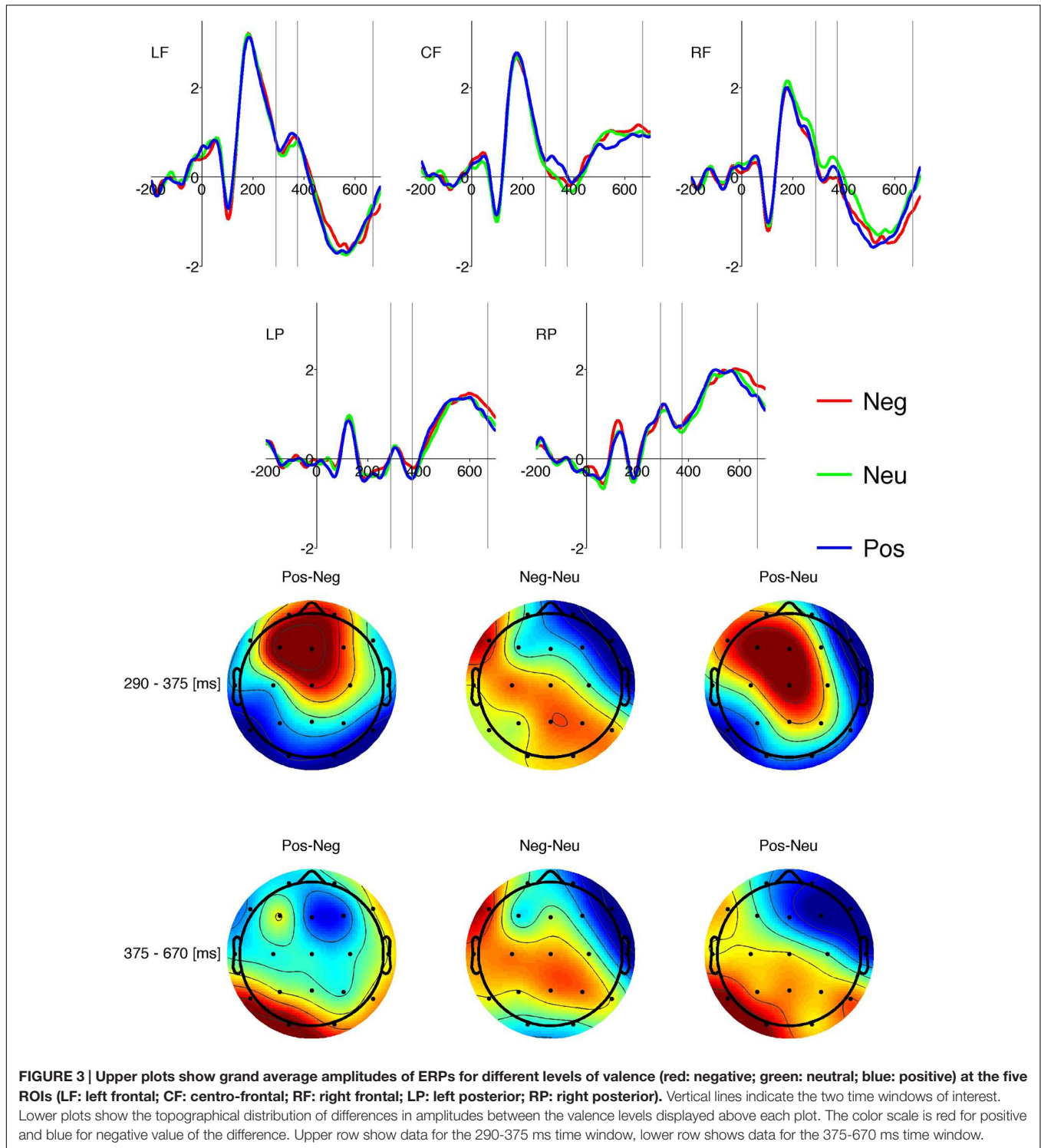
There was an interaction between valence and origin for the right-frontal ROI [ $F(4,124) = 4.43$ ,  $p < 0.003$ ]. *Post hoc* analysis showed that Neu\_0 stimuli elicited more positive amplitudes than Neg\_A stimuli [ $t(31) = 5.17$ ,  $p < 0.0005$ ], Pos\_0 stimuli [ $t(31) = 3.81$ ,  $p < 0.03$ ], Neg\_R stimuli [ $t(31) = 3.64$ ,  $p < 0.04$ ] and Neu\_R stimuli [ $t(31) = 3.53$ ,  $p < 0.05$ ]. The amplitudes for each condition are given in Appendix 1 (Table A4).

#### 375-670 time window (corresponding with LPC)

There was an interaction between origin and ROI [ $F(8,248) = 2.46$ ,  $p < 0.014$ ]. There was a main effect of origin at the right frontal [ $F(2,62) = 5.13$ ,  $p < 0.01$ ] and left parietal [ $F(2,62) = 6.39$ ,  $p < 0.003$ ] ROIs. **Figure 5** shows that responses in the left parietal and right frontal ROIs show similar origin-related changes although the polarity of the responses is reversed, which is the signature of a dipolar pattern. *Post hoc* paired sample *t*-tests showed that at the left posterior ROI responses to A and R stimuli were more positive than responses to 0 stimuli [ $t(95) = 3.27$ ,  $p < 0.005$  and  $t(95) = 2.80$ ,  $p < 0.012$  respectively]. At the right frontal ROI only the difference between A and 0 was significant [ $t(95) = 3.05$ ,  $p < 0.01$ ]. The higher error in amplitude data rendered the differences between 0 and R non-significant. This result is visualized in **Figure 5**.

## DISCUSSION

Our study provides an orthogonal comparison of the impact on word processing of two properties of words, valence and origin. We controlled variability in factors such as arousal, concreteness, frequency and word length. Our pseudo-word stimuli were generated carefully and were matched in length with the word stimuli.

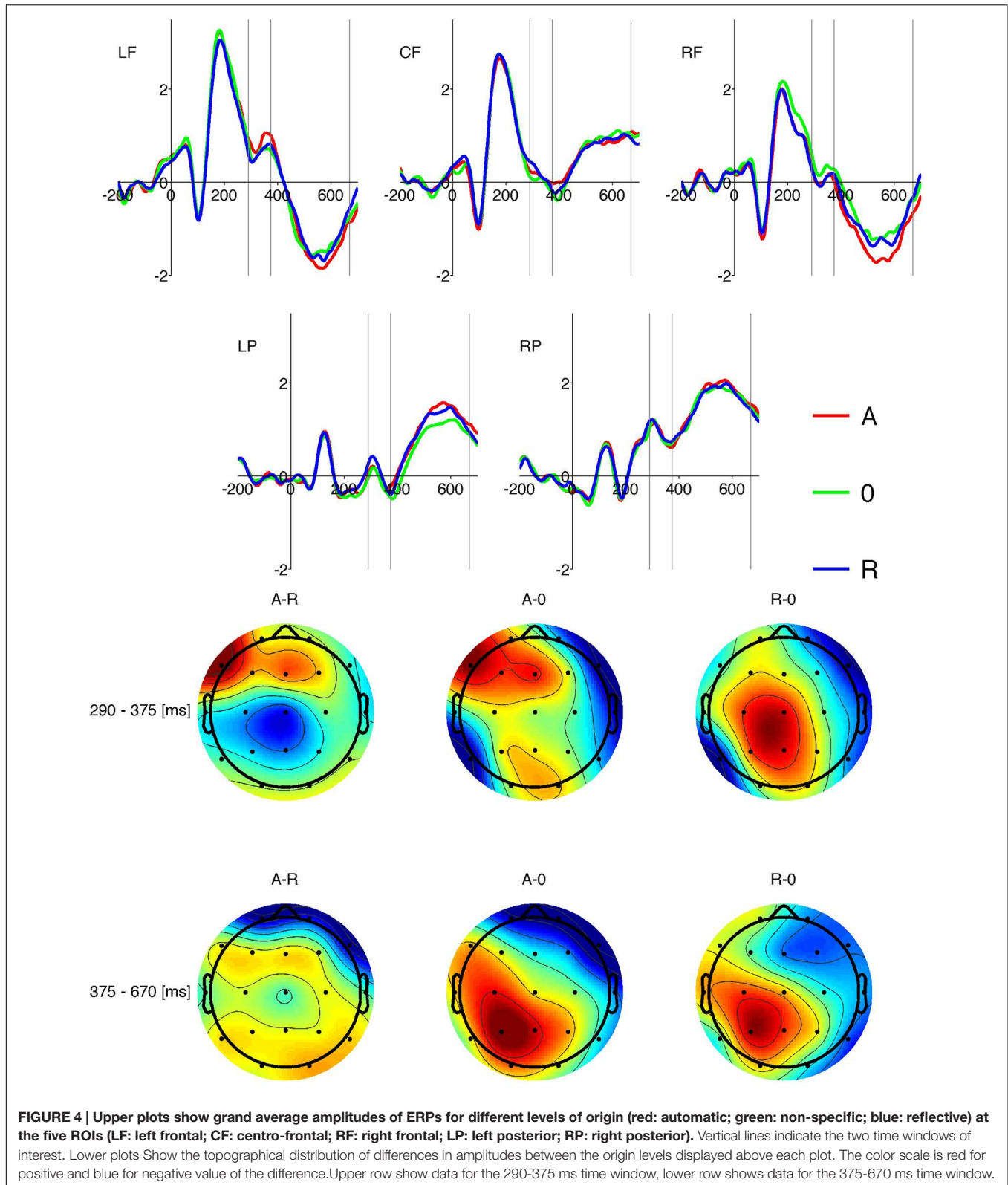


**FIGURE 3 |** Upper plots show grand average amplitudes of ERPs for different levels of valence (red: negative; green: neutral; blue: positive) at the five ROIs (LF: left frontal; CF: centro-frontal; RF: right frontal; LP: left posterior; RP: right posterior). Vertical lines indicate the two time windows of interest. Lower plots show the topographical distribution of differences in amplitudes between the valence levels displayed above each plot. The color scale is red for positive and blue for negative value of the difference. Upper row show data for the 290-375 ms time window, lower row shows data for the 375-670 ms time window.

### Behavioral Results

Response times were longer for pseudo-words than for words, which suggests that decision making was more difficult in the case of pronounceable but meaningless stimuli. Contrary, there was no effect of valence or origin on response times in the LDT. This is consistent with a study by Barber et al. (2013)

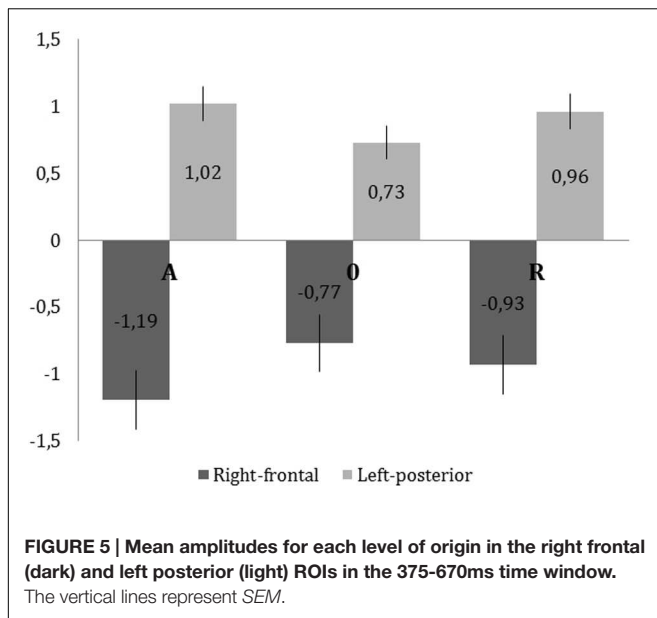
which found that concreteness but not valence affected response latencies, with abstract words eliciting faster responses than concrete words. Another study (Kanske and Kotz, 2007) also found that response latency was affected by concreteness, but not other word properties; however, in this study the opposite pattern was observed: concrete words elicited faster responses



**FIGURE 4 |** Upper plots show grand average amplitudes of ERPs for different levels of origin (red: automatic; green: non-specific; blue: reflective) at the five ROIs (LF: left frontal; CF: centro-frontal; RF: right frontal; LP: left posterior; RP: right posterior). Vertical lines indicate the two time windows of interest. Lower plots Show the topographical distribution of differences in amplitudes between the origin levels displayed above each plot. The color scale is red for positive and blue for negative value of the difference. Upper row show data for the 290-375 ms time window, lower row shows data for the 375-670 ms time window.

than abstract ones. From that reason, our results support the claim that origin and concreteness are distinct constructs (c.f. Introduction). Positive words were more likely to be classified

correctly than negative and neutral words. Furthermore, words of no specific origin (control condition) were less likely to be classified correctly than automatic and reflective words. Further



analysis revealed that this effect was due to poor classification of one stimulus category, namely negative words of no specific origin (c.f. **Table 2**). Barber et al. (2013) found a trend toward more accurate classification of concrete stimuli; concrete verbs were discriminated slightly more accurately than abstract words and pseudo-words. All electrophysiological data analyses were conducted using only data from accurate trials, thus we cannot attribute the obtained results strictly to the error rates of responses. Nevertheless those differences should be taken into account when interpreting the results.

## Differences between Words and Pseudo-Words

The amplitude of responses to words and pseudo-words differed in three time windows, namely the 225-290 ms, 290-375 ms, and 375-670 ms windows. This effect was modulated by electrode position. In the case of the 225-290 time window, differences were detectable in the centro-frontal ROI and words elicited larger - more positive - amplitude responses than pseudo-words. More research is needed to replicate this effect and evaluate conclusions that can be drawn from it. In the 290-375 ms time range identified as an N400 or FN400 component, the pseudo-words generated larger - more negative - responses than words in all frontal ROIs, whereas the opposite pattern of results (words elicit larger negative responses) was observed in the left parietal ROI.

In the 375-670 ms time range identified as a LPC component, words generated larger amplitude responses than pseudo-words at the centro-frontal, left parietal and right parietal ROIs; in the left and right frontal ROIs the reverse pattern was observed. The absence of differences between words and pseudo-words in the early ERP component (110-225 ms: N200) suggests that there were no orthographical or other formal differences between the word and pseudo-word stimuli used in our study (Nobre et al., 1994). The FN400 component responses suggest some kind of

a surprise associated with the processing of pseudo-words of no semantic meaning in the frontal regions of the brain (Bentin et al., 1999), which may be related to the greater reading difficulty of pseudo-words; this suggestion is supported by the reaction time data. Finally, the LPC effect suggests that meaningful stimuli elicit deeper processing of word meaning (Citron, 2012). The results of comparisons between pseudo-words and words suggest the paradigm used was valid and confirms the stages of visual word processing identified in earlier studies (Bentin et al., 1999).

## Differences Related to Valence and Origin

### The FN400 Component

There was a main effect of valence at the centro-frontal ROI for amplitudes in the 290-375 ms time range. Positive words elicited more positive amplitude than negative and neutral words. This effect was detectable at frontal ROIs, which is consistent with the word versus pseudo-word findings in this time range and suggests that activity in this time range should be interpreted as an FN400 component (Curran, 2000; Voss and Paller, 2007; Kutas and Federmeier, 2011) rather than an N400 component (Bentin et al., 1999; Kanske and Kotz, 2007; Kutas and Federmeier, 2011) or an EPN as in some LDT studies (Bayer et al., 2012). The FN400 component is thought to be related to semantic processing, especially the link between stimulus and meaning (Kutas and Federmeier, 2011). One would expect to detect valence effects on this component, because positive, negative and neutral valenced words are grouped semantically in the mind (Kanske and Kotz, 2007). Previous findings on concreteness are not consistent. Kanske and Kotz (2007) found that concrete nouns elicited a larger N400 response than abstract nouns but Palazova et al. (2013) reported no concreteness effect for verbs in the EPN time range. Given that valence is an intuitive dimension and should be processed sooner and more easily than other dimensions such as concreteness (Bayer et al., 2012; Palazova et al., 2013), we expected that FN400 amplitude would be more strongly influenced by valence than by other word properties. The data confirmed this prediction: we found that ERP responses varied according to valence.

In the right frontal ROI there was an interaction between valence and origin. Since Neu\_0 words elicited larger amplitude responses than the four other valence-origin combinations it is worth inspecting word properties. All stimulus groups were matched for concreteness, arousing properties and frequency of appearance in language, but there were some group differences in word length. Words in the Neu\_0 group were on average about one letter shorter than those in other word groups (c.f. **Table 1**; Appendix 1). Shorter stimuli are expected to be processed in an easier way. In current study paradigm we may assume that in the right frontal ROI response amplitude is negatively associated with stimulus length (c.f. LPC discussion).

### The LPC Component

The LPC component was found to be sensitive to concreteness in the case of both nouns (Kanske and Kotz, 2007) and verbs (Palazova et al., 2013). Abstract nouns and verbs elicited larger LPC amplitude responses than concrete nouns and verbs. In our

study the amplitude in the LPC time range differed according to the origin of the word in two ROIs, namely the right frontal and left parietal regions. In the right frontal ROI automatic words elicited more negative amplitude responses than 0 words whereas in the left parietal ROI both automatic and reflective stimuli elicited more positive amplitude response than 0 words. The scalp location and pattern of differences in both ROIs suggest activation of a dipolar source; amplitudes at the same locations were affected by word origin in a task requiring explicit emotional processing (Imbir et al., 2015a). Alternatively these results could be interpreted as effects of different, temporally correlated processes; the right frontal ROI effect could be related to differences in stimulus length and the left parietal effect could be related mostly to differences in origin of words.

Comparing the ERP results and behavioral data shows that words of no specific origin (0 words) were classified less accurately than words with a specific origin (automatic and reflective words). This suggests that specified origin of an affective component of word make decisions easier (after controlling for potential effects of frequency of appearance, concreteness and arousing properties). Since opposite patterns were observed in the two described above ROIs it is hard to explain the results in behavioral terms. It is also important to consider the group difference in word length in our selection of stimuli. Automatic words were longer than 0 words (c.f. Linguistic materials properties and **Table 1**). In line with FN400 component findings for the right frontal ROI we predicted that shorter words would elicit larger amplitude responses than longer words. In fact this pattern is replicated in LPC time range in the same ROI. It should be noted that there was no difference in length between reflective and 0 words, just as there was no difference in the amplitude of responses to these stimulus types in the right frontal ROI. We can therefore conclude that responses in the right frontal ROI are sensitive to stimulus length.

In our previous studies using the odd-ball task to assess effects of the emotionality of words (emotional versus neutral valence character assessments; Imbir et al., 2015a), we found that the LPC response was influenced by the origin of emotion. Raw ERP amplitude was higher for words of reflective origin than words of automatic origin, but this effect could have been due to differences in the concreteness of the stimuli. ICA showed an independent component with a dipolar topography in left posterior locations for which amplitude was different only in the case of valenced (not differing from concreteness), but not neutral words (differing from concreteness). The amplitude of this component was larger for words of automatic origin than those of reflective origin.

Current study, which used a different task and different verbal stimuli, has confirmed our previous findings about location of region engaged in processing. It is worth highlighting that we found some differences in pattern of results. Words of automatic origin elicited more positive LPC responses than words of unspecified origin (this study) or words of reflective origin (previous study). This difference could be attributed to the judgment participants were required to make about stimuli in the two studies. In the earlier study (Imbir et al., 2015a)

participants had to decide whether a word was emotional or neutral (explicit lexical processing of meaning), whereas in this study they had to classify stimuli as words or pseudo-words (implicit lexical processing). These tasks depend on underlying mental processes which differ with respect to depth and profile of analysis. Origin is an emotional property referring to whether processing engages the AES or RES. When we asked participants about the emotional quality of stimuli, the AES was associated with more crucial experiences (such as threats to life); these should be deeper and produce a higher amplitude than RES experiences. When we asked participants about lexical quality, the differences between systems could be less salient, but still the AES should attract more attention.

## CONCLUSION

We have presented data relevant to how lexical processing of words in a LDT is influenced by the valence and origin of stimuli. Valence influenced the amplitude of the FN400 component, whereas origin influenced the LPC. The study was designed carefully to avoid any potentially confounding factors biasing the results; we controlled for potential effects of concreteness, arousal, frequency of appearance and stimulus length in a factorial design. We found no reaction latencies differences across conditions, but we did find accuracy effects. Effects on the FN400 and LPC components in the right frontal region could be attributed to differences in task difficulty and stimulus length. Our findings on the effects of origin are consistent with an earlier study (Imbir et al., 2015a) indicating that left parietal regions are engaged in processing of stimulus origin. This suggests that the origin of emotional states is one of the factors that modulate late stages of word processing. The results of this study are important for understanding the role of complexity represented in stimuli evoking automatic or reflective originated emotional responses. We showed that this complexity is affective in character and not related to the concreteness of stimuli. What is more, this origin related complexity influences not only explicit, but also implicit processing of semantic stimuli.

## AUTHOR CONTRIBUTIONS

All authors contributed to final version of the manuscript. Theoretical proposition: KI; Design: KI, JZ; Method (words): KI; Method (EEG measures) JZ, TS; Experimental procedure programming: TS, JZ; Experiment execution: TS, JZ; Statistical analyses: JZ, KI, TS; Results description: JZ; Results discussion: KI; Figures: JZ, TS, KI; Tables: JZ, KI.

## FUNDING

The project was funded by the National Science Center on the basis of decision DEC: DEC-2013/09/B/HS6/00303.

## ACKNOWLEDGMENTS

We would like to express our thanks to Alicja Brzozowska for participation in data collection and technical assistance.

## REFERENCES

- Barber, H. A., and Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Res. Rev.* 53, 98–123. doi: 10.1016/j.brainresrev.2006.07.002
- Barber, H. A., Otten, L. J., Kousta, S. T., and Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain Lang.* 125, 47–53. doi: 10.1016/j.bandl.2013.01.005
- Bayer, M., Sommer, W., and Schacht, A. (2012). Font size matters—emotion and attention in cortical responses to written words. *PLoS ONE* 7:e36042. doi: 10.1371/journal.pone.0036042
- Bentin, S., Kutas, M., and Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology* 30, 161–169. doi: 10.1111/j.1469-8986.1993.tb01729.x
- Bentin, S., Mouchetant-Rostaing, Y., Giard, M., Echallier, J., and Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J. Cogn. Neurosci.* 11, 235–260. doi: 10.1162/089892999563373
- Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain Lang.* 122, 211–226. doi: 10.1016/j.bandl.2011.12.007
- Citron, F. M. M., Oberecker, R., Friederici, A. D., and Mueller, J. L. (2011). Mass counts: ERP correlates of non-adjacent dependency learning under different exposure conditions. *Neurosci. Lett.* 487, 282–286. doi: 10.1016/j.neulet.2010.10.038
- Crowne, D. P. (2007). *Personality Theory*. Don Mills, ON: Oxford University Press.
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Mem. Cognit.* 28, 923–938. doi: 10.3758/BF03209340
- Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., and Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol. Psychol.* 52, 95–111. doi: 10.1016/S0301-0511(99)00044-7
- Damasio, A. R. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. New York, NY: Pantheon Books.
- Darlow, A. L., and Sloman, S. A. (2010). Two systems of reasoning: architecture and relation to emotion. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 382–392. doi: 10.1002/wcs.34
- 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
- Dillon, D. G., Cooper, J. J., Woldorff, M. G., and LaBar, K. S. (2006). Dissociation of event-related potentials indexing arousal and semantic cohesion during emotional word encoding. *Brain Cogn.* 62, 43–57. doi: 10.1016/j.bandc.2006.03.008
- Epstein, S. (2003). "Cognitive-experiential self-theory of personality," in *Handbook of Psychology*, Vol. 5, *Personality and Social Psychology*, eds T. Millon and M. J. Lerner (Hoboken, NJ: John Wiley & Sons, Inc.), 159–184.
- Fischler, I., and Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. *Prog. Brain Res.* 156, 185–204. doi: 10.1016/S0079-6123(06)56009-1
- Gawronski, B., and Creighton, L. A. (2013). "Dual-process theories," in *The Oxford handbook of Social Cognition*, ed. D. E. Carlston (New York, NY: Oxford University Press), 282–312.
- Herbert, C., Herbert, B. M., Ethofer, T., and Pauli, P. (2011). His or mine? The time course of self-other discrimination in emotion processing. *Soc. Neurosci.* 6, 277–288. doi: 10.1080/17470919.2010.523543
- Herbert, C., Junghofer, M., and Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology* 45, 487–498. doi: 10.1111/j.1469-8986.2007.00638.x
- Herbert, C., Pauli, P., and Herbert, B. M. (2010). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Soc. Cogn. Affect. Neurosci.* 2, 303–312. doi: 10.1093/scan/nsq082
- Imbir, K. (2013). Origins and sources of emotion as factors that modulate the scope of attention. *Ann. Psychol.* 16, 109–132.
- Imbir, K. (2015). Affective norms for 1,586 Polish words (ANPW): Duality of mind Approach. *Behav. Res. Methods* 47, 860–870. doi: 10.3758/s13428-014-0509-4
- Imbir, K. (2016). Subjective significance shapes arousal effects on modified stroop task performance: a duality of activation mechanisms account. *Front. Psychol.* 7:75. doi: 10.3389/fpsyg.2016.00075
- Imbir, K., and Jarymowicz, M. (2013). The Effect of Automatic vs. reflective emotions on cognitive control in antisaccade tasks and the emotional stroop test. *Pol. Psychol. Bull.* 44, 137–146. doi: 10.2478/ppb-2013-0016
- Imbir, K., Jarymowicz, M., Spustek, T., Kuś, R., and Żygierewicz, J. (2015a). Origin of emotion effects on ERP correlates of emotional word processing: the emotion duality approach. *PLoS One* 10:e0126129. doi: 10.1371/journal.pone.0126129
- Imbir, K., Spustek, T., and Żygierewicz, J. (2015b). Polish pseudo-words list: dataset of 3023 stimuli with competent judges' ratings. *Front. Psychol.* 6:1395. doi: 10.3389/fpsyg.2015.01395
- Jarymowicz, M., and Imbir, K. (2015). Toward a human emotions taxonomy (based on their automatic vs. reflective origin). *Emot. Rev.* 7, 183–188. doi: 10.1177/1754073914555923
- Kagan, J. (2007). *What is Emotion? History, Measures, and Meanings*. New Haven, CT: Yale University Press.
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *Am. Psychol.* 58, 697–720. doi: 10.1037/0003-066X.58.9.697
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.
- Kanske, P., and Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res.* 1148, 138–148. doi: 10.1016/j.brainres.2007.02.044
- Kazojć, J. (2011). *Słownik Frekwencyjny Języka Polskiego [Polish Language Dictionary of Attendance]*. Available at: <http://www.slovníki.org.pl/lista-frekwencyjna.pdf>
- Kissler, J., Herbert, C., Peyk, P., and Junghofer, M. (2007). Buzzwords early cortical responses to emotional words during reading. *Psychol. Sci.* 18, 475–480. doi: 10.1111/j.1467-9280.2007.01924.x
- Kissler, J., Herbert, C., Winkler, I., and Junghofer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biol. Psychol.* 80, 75–83. doi: 10.1016/j.biopsycho.2008.03.004
- 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., Lindamood, T. E., and Hillyard, S. A. (1984). "Word expectancy and event-related brain potentials during sentence processing," in *Preparatory States and Processes*, eds S. Kornblum and J. Requin (Hillsdale, NJ: Erlbaum), 217–238.
- Lehmann, D., and Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.* 48, 609–621. doi: 10.1016/0013-4694(80)90419-8
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: The MIT Press.
- Münste, T., Matzke, M., and Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *J. Cogn. Neurosci.* 9, 318–329. doi: 10.1162/jocn.1997.9.3.318
- Nobre, A. C., Allison, T., and McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature* 372, 260–263. doi: 10.1038/372260a0

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00271>

- Osgood, C. E., Suci, G., and Tannenbaum, P. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Palazova, M., Sommer, W., and Schacht, A. (2013). Interplay of emotional valence and concreteness in word processing: an event-related potential study with verbs. *Brain Lang.* 125, 264–271. doi: 10.1016/j.bandl.2013.02.008
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org>.
- Rastle, K. (2007). “Visual word recognition,” in *The Oxford Handbook of Psycholinguistics*, ed. M. G. Gaskell (Oxford: Oxford University Press), 71–88.
- Reykowski, J. (1989). “Dimensions of development of moral values,” in *Social and Moral Values*, eds N. Eisenberg, J. Reykowski, and E. Staub (Hillsdale, NJ: Erlbaum), 23–44.
- Rolls, E. T. (2000). Precise of the brain and emotion. *Behav. Brain Sci.* 23, 177–191. doi: 10.1017/S0140525X00432425
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145
- Schacht, A., and Sommer, W. (2009a). Emotions in word and face processing: early and late cortical responses. *Brain Cogn.* 69, 538–550. doi: 10.1016/j.bandc.2008.11.005
- Schacht, A., and Sommer, W. (2009b). Time course and task dependence of emotion effects in word processing. *Cogn. Affect. Behav. Neurosci.* 9, 28–43. doi: 10.3758/CABN.9.1.28
- Scherer, K. R. (2005). What are emotions? And how they can be measured? *Soc. Sci. Inform.* 44, 693–727. doi: 10.1177/0539018405058216
- Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., and Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology* 37, 257–261. doi: 10.1111/1469-8986.3720257
- Schupp, H. T., Cuthbert, B., Bradley, M., Hillman, C., Hamm, A., and Lang, P. (2004). Brain processes in emotional perception: motivated attention. *Cogn. Emot.* 18, 593–611. doi: 10.1080/02699930341000239
- Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topogr.* 3, 137–141. doi: 10.1007/BF01128870
- Strack, F., and Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* 8, 220–247. doi: 10.1207/s15327957pspr0803\_1
- Strack, F., and Deutsch, R. (2014). “The reflective–impulsive model,” in *Dual-Process Theories of the Social Mind*, eds J. W. Sherman, B. Gawronski, and Y. Trope (New York, NY: Guilford Press), 92–104.
- Thomas, D. G., Grice, J. W., Najm-Briscoe, R. G., and Williams Miller, J. (2004). The influence of unequal numbers of trials on comparisons of average event-related potentials. *Dev. Neuropsychol.* 26, 753–774. doi: 10.1207/s15326942dn2603\_6
- Voss, J. L., and Paller, K. A. (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. *Learn. Mem.* 14, 259–267. doi: 10.1101/lm.529807

**Conflict of Interest Statement:** 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.

The reviewer DG and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review

Copyright © 2016 Imbir, Spustek and Żygierewicz. 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) or licensor 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.



# Processing Code-Switching in Algerian Bilinguals: Effects of Language Use and Semantic Expectancy

Souad Kheder\* and Edith Kaan

Department of Linguistics, University of Florida, Gainesville, FL, USA

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Maria Garraffa,  
Heriot-Watt University, UK  
Roberto Ramirez Heredia,  
Texas A&M International University,  
USA

### \*Correspondence:

Souad Kheder  
skheder@ufl.edu

### Specialty section:

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

**Received:** 20 September 2015

**Accepted:** 08 February 2016

**Published:** 01 March 2016

### Citation:

Kheder S and Kaan E (2016)  
Processing Code-Switching  
in Algerian Bilinguals: Effects  
of Language Use and Semantic  
Expectancy. *Front. Psychol.* 7:248.  
doi: 10.3389/fpsyg.2016.00248

Using a cross-modal naming paradigm this study investigated the effect of sentence constraint and language use on the expectancy of a language switch during listening comprehension. Sixty-five Algerian bilinguals who habitually code-switch between Algerian Arabic and French (AA-FR) but not between Standard Arabic and French (SA-FR) listened to sentence fragments and named a visually presented French target NP out loud. Participants' speech onset times were recorded. The sentence context was either highly semantically constraining toward the French NP or not. The language of the sentence context was either in Algerian Arabic or in Standard Arabic, but the target NP was always in French, thus creating two code-switching contexts: a typical and recurrent code-switching context (AA-FR) and a non-typical code-switching context (SA-FR). Results revealed a semantic constraint effect indicating that the French switches were easier to process in the high compared to the low-constraint context. In addition, the effect size of semantic constraint was significant in the more typical code-switching context (AA-FR) suggesting that language use influences the processing of switching between languages. The effect of semantic constraint was also modulated by code-switching habits and the proficiency of L2 French. Semantic constraint was reduced in bilinguals who frequently code-switch and in bilinguals with high proficiency in French. Results are discussed with regards to the bilingual interactive activation model (Dijkstra and Van Heuven, 2002) and the control process model of code-switching (Green and Wei, 2014).

**Keywords:** code-switching, semantic constraint, base language, language expectancy, processing, listening

## INTRODUCTION

Some bilingual speakers may daily interact in a context where similar speakers use both languages in the same conversation or even in the same utterance of speech. This type of interaction is usually known as code-switching (CS). During code-switching, bilinguals may be listening to an utterance that starts in their first language (L1) but may or may not end in that same language depending on the speaker's speech planning (e.g., Kroll and Gollan, 2014). How does a bilingual listener integrate an item from their second language (L2) while listening to L1? While the choice of code-switching is made by the speaker, this choice may still impact the listener. Listeners are active recipients who constantly make inferences about the speakers' intentions, and may develop models on how they may respond (Gross, 2000; Green and Abutalebi, 2013). We know from studies which tested bilingual speakers during reading in L1 or in L2 that the semantic



and syntactic context of a sentence speeds up word recognition, assumingly through reducing the number of the activated candidates. Words embedded in a semantic context that is highly constraining are processed faster than words embedded in a semantic context that is neutral (e.g., Schwartz and Kroll, 2006; Duyck et al., 2007; van Hell and de Groot, 2008; Van Assche et al., 2009, 2011; Titone et al., 2011). Studies which explored sentential context influence during switching mainly concerned reading from L2 to L1. While code-switching is primarily conversational where bilinguals may either be speakers (production) or listeners (comprehension), relatively fewer studies focused on listening comprehension. Furthermore, previous studies often failed to give details concerning the switching behavior of their participants in their daily lives. Hence, the populations tested previously may have included speakers who did not switch frequently in natural conversations. The daily use of the bilingual languages and their different interactional contexts, may have a role in shaping the use of sentence constraints. In multilinguals, code switching may be more frequent in one language combination than another. Thus, it is important to explore sentential effects during code-switching taking into account the daily language use of the bilinguals (e.g., Green, 2011). The main objective of the current study is to determine to what extent (1) the language that precedes a code-switch, and (2) the semantic context affect the expectancy of a language switch during listening to L1, and to see if this effect is modulated by (3) code-switching habits and (4) the L2 proficiency of the bilinguals.

## Effects of Semantic Constraint

Altarriba et al. (1996) explored the effect of semantic context constraints in mixed-language sentences in Spanish–English bilinguals during reading for comprehension. Their goal was to determine whether sentence context effects can extend to the determination of the lexical features. Participants read Spanish and English target words inserted in English low and high constraint sentences while fixations were measured with an eye tracker. The results revealed that high frequency Spanish words produced slower naming times and longer fixation times when they appeared in high constraint English sentences but not in a low constraint context. This suggests that in the highly constrained context the readers generated semantic and lexical features of the upcoming words in the English context. When the target word appeared in Spanish, not the language of the context, the expectations regarding the lexical features were not met. The readers expected a word form with a specific meaning, suggesting that semantic context can selectively activate a word in one language.

The effect of sentence context in code-switched sentences was also found in studies using event-related potentials (ERPs; e.g., Moreno et al., 2002; Proverbio et al., 2004). Moreno et al. (2002) explored the ERPs to a language switch when English–Spanish bilinguals read both moderately constraining and highly constraining English sentences. The sentences appeared in three conditions: either ending with expected English words, ending with their Spanish translations (code-switches), or ending with English synonyms (lexical switches). Code-switches elicited a

large posterior positivity in both context types usually associated with the processing of an unexpected or improbable event. The authors suggested that the Spanish words acted like improbable events probably because they occurred in written text and were from English to Spanish while natural code-switching occurs mostly in speaking and is from Spanish to English. Moreno et al. (2002), thus, suggested that more natural code-switches, would reduce the positivity making the code-switch less improbable. In Proverbio et al. (2004), Italian professional interpreters evaluated whether a sentence-final word made sense with the rest of the sentence in a unilingual or a mixed/code-switched condition. Sentences ended in an unexpected final word that created a semantic incongruence or in a highly probable and congruent word. Even when the switch was entirely predictable the mixed conditions revealed processing cost. A larger N400, associated with semantic integration, was reported for the final words in the mixed compared to the unmixed sentences. Although the participants in these studies were highly proficient, the authors did not report to what extent these participants mixed languages in daily conversation.

In a cross-modal naming (CMN) paradigm (i.e., naming a visual word in an auditory context) Hernández et al. (1996) explored the effect of expectancy and predictability in sentential priming within and between languages during listening. Spanish–English bilinguals heard sentences and named a word that appeared on the screen at a certain point during the presentation of the sentence. The target word was either related or unrelated to a critical word in the sentence and was either presented immediately or delayed. In the mixed condition (English–Spanish or Spanish–English sentences), facilitation occurred only under the delayed naming condition. The results showed that cross-language priming, that is semantic and lexical facilitation, appears when participants expect the language of the target word, when they have sufficient time to generate a response, that is, to access and integrate the target word in the sentence context, or both. In addition, the results also showed that priming was larger for English than for Spanish suggesting that priming is more robust for the language that bilinguals are using most often in their everyday lives. Hernández et al. (1996) reported that the participants in this study were native speakers of Spanish who were immersed in an English environment since they started school and thus may have been English-dominant. While it is possible that English dominated in the participants' speech, they may or may not be from a code-switching community.

Cieślicka and Heredia (2015) used CMN to investigate the effect of context and cross-language priming on lexical access in Spanish–English bilinguals. Participants listened to sentences in Spanish or English that contained a prime (e.g., *war*) and then named a target word either in Spanish or in English. The target was either related (e.g., *peace*) or unrelated (e.g., *boca*) to the prime. In addition, sentence context was either biasing or not toward the target prime. The targets were presented at the offset of the prime to examine the activation of L1 after the processing of L2 and vice versa. When the context language was Spanish, context manipulation did not modulate priming effect between within and cross-language conditions, but it affected the overall processing of the target words in the Spanish–English conditions.

The targets in the biased context were faster to name than those in the unbiased context. However, when the context language was English, cross-language priming effect was greater than within-language priming effect and biased context slowed down the naming of the target words. An important finding in this study was the language asymmetry in lexical access. Participants were faster to respond to words in the L2 when the preceding context was in the L1 than vice versa. The interpretation Cieřlicka and Heredia (2015) suggested was that with increased proficiency in L2, the ease with which bilinguals access words in their languages depends on language usage.

## Code Switching Habits and Effect of Base-Language

The overall results from the above studies suggest that semantic context as well as the language of the preceding sentence affect the processing of upcoming words in bilingual speakers. However, different bilinguals may have different linguistic experiences that have prepared them as speakers/listeners to use one or the other language separately in one context, or even use both in the same utterance in another context (e.g., Grosjean, 2008). Meuter (2009) sees the ability to select among these options as a sign of highly proficient language use and goal-directed behavior but at the same time dictated by communicative contexts. Specifically, in a code-switching context, bilinguals must be ready to integrate (comprehension) and respond (production) in any language, while being sensitive to the cues of each language such as accent and language-specific morpho-syntactic patterns. Thus, processing a language switch may depend greatly on language use and switching habits. For instance in an eye tracking study, Valdés Kroff et al. (in press) found that the masculine Spanish article *el* functioned as a default article with English nouns in Spanish–English code-switching resulting in competitor effects due to phonological competition over the feminine article *la*. When the Spanish article + English noun code-switches were tested with bilinguals who were not code-switchers (Valdés Kroff et al., 2011), a processing delay was found for both *el* and *la* with English noun combinations. Results from these studies suggest that bilinguals who are exposed to code-switching showed a pattern of processing code-switches that was different from the bilinguals who were not exposed to code-switching. The latter group suffered greater processing costs when they encountered code-switches.

There has been an increased focus on language use and switching habits of the bilinguals and their effect on switching control processing. Green and Wei (2014) have proposed a control process model based on the adaptive control hypothesis (Green and Abutalebi, 2013). The adaptive control hypothesis proposes a language control system where processes vary with different interactional contexts. Three different interactional contexts are identified in a bilingual setting (Green and Abutalebi, 2013; Green and Wei, 2014). In a “single language context” one language is used in one context (e.g., at home) and the other used in another context (e.g., at school/work). In a “dual language context,” both languages may be used in the same environment but with different speakers. Finally, in a “dense

code-switching context” the bilinguals habitually switch between their languages within the same utterance and adapt words from one language to fit within the structure of the other. Therefore, in the dual context there is higher demand on processes that control interference in order to minimize inappropriate switching. The language schemas are in a competitive relationship and control alternates between the schemas of the different languages. In the dense code-switching context, there is increased demand on control processes that allow alternative forms. Hence, language task schemas cooperate to allow alternative forms depending on their appropriateness in the given context. The hypothesis states that different contexts of language use may shape the adaptation of control processes, i.e., a bilingual’s experience using both languages will shape the specific cognitive mechanisms that more generally support bilingual language control. The control system adapts to the demands of the different interactions to avoid or reduce the interactional cost that may arise when bilinguals from different interactional contexts converse. For instance, when bilinguals who code-switch feel the necessity to avoid code-switching in a given interaction due to inappropriateness, they face an interactional cost. This is because this interaction imposes on them to remain in one language and block interference from the other. To control for interference they have to engage some control processes in which they are not well-trained and this incurs extra processing load.

When dense code-switching is not a common language practice of the bilinguals, it is likely that encountering a lexical switch within the same utterance imposes higher cognitive processing demands on control processes in which these bilinguals are not well-trained. Similarly, when bilinguals from a dense code-switching context encounter code-switches that do not allow alternative forms (adaptation), they are forced to use competitive control processes not typical of their language processing. We may speculate that the conflicting results reported in previous studies concerning the ease of code-switching may be due to incongruities between the participants’ habitual interactional contexts and the type of code-switched material on which they have been tested.

Bilinguals who code-switch may also differ in respect to the direction of code-switching, that is, the language they switch from or the “base language.” Many code-switching studies examined switching from L2 to L1. Although this switching pattern is attested among bilingual communities, it is less common than switching from L1 to L2 (e.g., Moreno et al., 2002). For instance Spanish–English bilinguals in Texas code-switch more to their L2 English when they communicate in their L1 Spanish than they do to L1 Spanish when they communicate in L2 English (Heredia and Altarriba, 2001). Additionally, in multilingual communities in which speakers may use more than two languages, switching may involve typically one pair of the languages but not the other. Since switching between languages is conversational, it is likely that the languages that are used in everyday conversations are the ones which are involved in code-switching. In the present study we examined the effect of semantic constraint and the language preceding the switch on processing code-switching by comparing two types of mixed-language sentences, one which typically occurs in everyday conversations and the other which

does not. We explored Algerian bilinguals who belong to a community where code-switching is a well-established way of communication (e.g., Boumans and Caubet, 2000) but who differ in the frequency of daily code-switching.

## The Current Study

The current experiment examined the effect of context language (base language), semantic constraints and language use on the expectancy of a language switch during listening comprehension. The habit and frequency of switching between a pair of languages rather than another may affect lexical expectancy and switching licensing. Code-switching between Algerian Arabic (AA) and French (FR) is conversational and frequent among some Algerian bilinguals but not code-switching between Standard Arabic (SA) and French. One of the possible reasons for this distribution is that although Algerians are introduced to Standard Arabic from the time they start school, and sometimes earlier, it is considered a school language used to study text books and get knowledge. Standard Arabic is never heard in conversations in the street or even in classrooms between students themselves. It is typical, however, to hear Algerian bilinguals speak Algerian Arabic and include French switches of varying length and morphological adaptation to Algerian Arabic structure (Boumans and Caubet, 2000; Bentahila and Davies, 2001, 2002).

We compared Algerian Arabic-French (AA-FR) code-switching to Standard Arabic-French (SA-FR) code-switching to investigate whether a language as a whole (base language) plays the role of a cue in expecting a language switch. In other words, when switching between a pair of languages that is typically attested in natural code-switching (AA-FR), Algerian bilinguals may expect a switch to French when they hear Algerian Arabic. However, when switching between a pair that is not so typical (SA-FR) Algerian bilinguals may not expect a language switch. In the latter case, code-switches may be harder to process and integrate with the preceding context than in the former case.

The second goal of this study was to examine whether the semantic constraints of the sentence context affects the expectancy of a language switch, and to compare the semantic constraints effect when switching between languages is typical (AA-FR) and when switching between languages is not typical (SA-FR). A high constraint context provides semantic cues that bias toward a specific lexical item and possibly its language (e.g., Altarriba et al., 1996; DeLong et al., 2005). On the other hand, a switch may be unexpected and hard to process even when it is highly predictable (Moreno et al., 2002; Proverbio et al., 2004). Accordingly, if a switch is unexpected in a more probable context (AA-FR) code switching, it should be more unexpected in a less probable context (SA-FR). Results from another experiment on Algerian bilinguals (Kheder et al., in preparation) that examined switch costs in high and low constraint contexts revealed semantic effects, that is, facilitation effect in high compared to low-constraint contexts regardless of word frequency, in both unilingual French context and AA-FR code-switching (as opposed to the Altarriba et al., 1996, in which context effects disappeared in high-constraint context for frequent target words). Based on the above, we predicted a language switch to be more expected in the high-constraint context than in the

low-constraint context. In addition, we expected the effect of semantic constraint to be larger when switching from Algerian Arabic base language than when switching from Standard Arabic.

The third goal was to see whether language expectancy in code-switching was affected by the bilingual's switching habits. Bilinguals who code-switch differ in their habits of using their languages. In a dense code-switching environment, bilinguals may interact in a context in which they switch languages between turns and sentences, or switch languages within the same utterance and tend to adapt words from one language to fit within the structure of the other (e.g., Green, 2011). They may also differ in the frequency and daily use of code-switching (Rodríguez-Fornells et al., 2011). The adaptive control hypothesis (Green and Abutalebi, 2013; Green and Wei, 2014) assumes that task schemas do not compete in a dense code-switching context, but cooperate to tolerate alternative forms depending on how much they fit within the given context. However, when dense code-switching is not a common language practice of the bilinguals between a pair of languages that they speak, it is likely that encountering a lexical switch within the same utterance is unexpected and imposes higher cognitive processing demands. Difference also exist among the Algerian bilinguals who interact in AA-FR dense code-switching: there are those who are heavy code-switchers (frequently code-switch) and others who are light code-switchers (not frequently code-switch). However, all these bilinguals do share the fact that they do not use Standard Arabic in everyday interactions and thus, SA-FR code-switching is not common practice for all of them. We expected larger effect of frequency of code-switching on language switch expectancy for light code-switchers than for heavy code-switchers. In particular, the difference between light and heavy switchers should be seen more in the AA-FR switches.

Finally, the study also sought to examine whether language expectancy in switching is modulated by French language proficiency. Proficiency may affect language activation in bilinguals with higher proficiency bilinguals showing more parallel activation than lower proficiency bilinguals (Blumenfeld and Marian, 2013), or showing a better control for L1 interference (Elston-Güttler et al., 2005). Proficient bilinguals may not need inhibition to produce words in one language only (e.g., Costa and Santesteban, 2004; Costa et al., 2006). Gollan and Ferreira (2009) noticed that balanced Spanish-English bilinguals switched languages more often than unbalanced bilinguals when they voluntarily switched languages in a naming task. The writers concluded that voluntary and cued language mixing became easier as proficiency increased because the more proficient bilinguals did not need to inhibit their dominant language in order to make the other language as much accessible. Language proficiency also affected accuracy rates in naming studies (e.g., Schwartz and Kroll, 2006) in which less proficient bilinguals had significantly higher naming error rates than the highly proficient bilinguals. In addition, proficiency affected the ability of the bilingual to stay in one language as needed and was seen to affect the word category that is most vulnerable to slips of the tongue (Poulisse, 2000), but also to affect the type of constituents in sentential code-switching among bilingual speakers (Backus, 1996; Myers-Scotton, 2006). Finally, language

proficiency was found to interfere with the effect of semantic constraint. Bilinguals with more proficiency in L2 showed reduced cognate facilitation in high-constraint sentences (e.g., Libben and Titone, 2009). The goal in the current study was therefore to see whether high proficient speakers of French differ from low proficient bilingual speakers of French in the expectancy of a language switch during listening to Algerian Arabic and Standard Arabic base languages.

To summarize, the research questions addressed in the current study are: (1) Is language expectancy in code-switching dependent on the base language? That is, does language expectancy differ between a typical code-switching (AA-FR) and a non-typical code-switching (SA-FR) context? (2) Do semantic constraints affect language expectancy in code-switching? (3) Is language expectancy dependent on the frequency of code-switching? (4) Does French L2 proficiency modulate the expectancy of language switching?

To answer these questions, we measured reaction times to the French NP code-switches using a CMN task (e.g., Hernández et al., 1996; Love et al., 2003). The CMN is an on-line method that is sensitive to sentential and lexical priming. In this task, the participants listen to a sentence, and at a particular point the sentence stops for a moment and a target word appears visually in the center of the computer screen. While the participants name the target word as fast and accurately as possible their reaction times are recorded. This task is similar to the Cross-Modal Lexical Priming task (CMLP; e.g., Li, 1996; Heredia and Blumentritt, 2002) with one critical difference. In the CMLP the flow of the sentence is not interrupted which makes it unlikely for the listener to engage in strategic processing (Heredia and Stewart, 2002). Thus, similarly to the CMLP task, the auditory presentation of the stimuli in the CMN used in the current study was not interrupted, and the visual target words appeared on the screen following the natural flow of the sentence without a time interval. The use of this task in this study allows us to obtain reaction times that reflect the processing of the context immediately as the sentence unfolds. Because the target words are presented at the offset of the last word that was heard, the effects observed on the participants' reaction times can be attributed to the participants' analysis of the sentence attained at the end of the auditory fragments.

The CMN has been demonstrated to measure what is active at certain moments in time during continuing processing of a sentence (e.g., Love et al., 2003). Its sensitivity to semantic as well as contextual effects have been well-documented (e.g., Tabossi, 1988, 1996; Hernández et al., 1996; Love et al., 2003). In particular, results from studies on code-switched sentences using CMN have shown that the paradigm is sensitive to sentential context (e.g., Hernández et al., 1996; Cieśllicka and Heredia, 2015). In contrast to Hernández et al. (1996), we presented sentences in a mixed rather than a blocked fashion because semantic and lexical facilitation in Hernández et al. (1996) occurred in the blocked condition when the bilinguals knew what language to expect. However, with a mixed block condition, it is unlikely that participants develop strategic responses since they are not aware of what comes next. In the latter case, any difference observed between the conditions would be the product

of sentential effects. In addition, priming effect in these studies depended on the language manipulation and was larger for the language of everyday interaction. Along with the advantages mentioned above, CMN is a good method to use in the current study for methodological reasons. The auditory presentation of the stimuli eliminates the issue of presenting Algerian Arabic, traditionally spoken only, in written script. Listening to the stimuli also avoid the visual appearance of both languages in the same sentence which is mostly stigmatized and considered ungrammatical by many bilinguals even those who code-switch. Since CMN is sensitive to what is active at points in time during the ongoing processing, it can test whether French is activated and available after listening to Algerian Arabic more than after listening to Standard Arabic.

Two factors were manipulated in this study: base language, that is, the language preceding the target word (Algerian Arabic or Standard Arabic) and semantic constraint of the context preceding the target word (High or low constrained contexts). The results in this study extend previous findings by exploring the effect of language use and frequency of code-switching. Algerian bilinguals listened to fragments of sentences either in Algerian Arabic or in Standard Arabic then immediately after named a target NP that was always in French. The target NP was thus heard in four different switching conditions: Algerian Arabic high-constraint context (AAH), Algerian Arabic low-constraint context (AAL), Standard Arabic high-constraint context (SAH), and Standard Arabic low-constraint context (SAL). Since all critical trials are code-switching trials, faster reaction times to the presented target words can be interpreted as ease of processing due to language switching expectation. Reaction times to French switches are compared in both base languages (AA and SA) and in both semantic constraint contexts (high and low). The following hypotheses and predictions are formed based on the research questions.

Concerning the first question, if the habit of switching between a certain pair of languages affects the expectancy of a language switching, there should be a base language effect. Participants should expect a switch to French when the base language is Algerian Arabic but not when the base language is Standard Arabic. This is because a switch to French is not typically expected when listening to Standard Arabic and because AA-FR code-switching is the default language switching in everyday conversation. Reaction times to French switches should be faster when Algerian Arabic is the base language than when Standard Arabic is the base language.

As regard to the second research question, if semantic context affects the expectation of a language switch, it is predicted that reaction times to switches in the high-constraint context should be faster than in the low-constraint context, and in particular after Algerian Arabic base language than after Standard Arabic base language. This is because the highly constraining context provides more semantic clues that help in predicting upcoming words, and previous studies showed that more predictable words are processed faster in naming (e.g., McClelland and O'Regan, 1981; Stanovich and West, 1981). In addition, expectations for a French continuation in a Standard Arabic context is weaker than in an Algerian Arabic context.

For the third research question, if language switch expectancy depends on the bilingual's recurrent switching habits, then reaction times for heavy code-switchers (those who frequently code-switch) should differ from light code-switchers (those who do not switch frequently). In addition, if heavy code-switchers are more experienced with dense code-switching contexts they should employ their cooperative control processes more than light code-switchers. In particular, we should see a difference in processing a switch depending on the extent of daily code-switching. In other words, bilinguals who code-switch more frequently should show more cooperative processes, which will be reflected in their reaction times. We also predict switching habits to interact with base language effect. Because AA-FR code-switching is more recurrent, anticipation of a language switch is more likely when Algerian Arabic is the base language. The difference between heavy code-switchers and light code-switchers should therefore be more apparent in AA-FR than in SA-FR code-switching. We finally predict switching habits to interact with semantic constraint. If language tasks schemas are in more cooperative mode for the heavy code-switchers compared to light code-switchers, then the effect of semantic constraint should differ between heavy and light code-switchers and more so in Algerian Arabic than in Standard Arabic base language.

For the last question, if French proficiency modulates the expectancy of language switching, then high proficiency bilinguals should be different from low proficiency bilinguals in processing the switch. If high proficiency bilinguals show more parallel activation for French, it is predicted that they should be faster overall than low proficient bilinguals and they should show a smaller effect for base language than low proficient bilinguals. Proficiency in French may also modulate the effect of semantic constraint. Highly proficient bilinguals in Libben and Titone (2009) showed reduced cognate facilitation in the high-constraint context which let them conclude that proficiency with semantic constraint can support language selectivity during the early stages of lexical access. If this is the case, high proficient bilinguals may show smaller semantic constraint effects than low proficient bilinguals.

## MATERIALS AND METHODS

### Ethical Approval

The current study was approved by the University of Florida Institutional Review Board (IRB) 02: Protocol #2014-U-0904.

### Participants

Sixty-five Algerian college students mostly from the National School of Computer Science and the National School of Polytechnics in Algiers participated in this experiment (mean age 22, range 18–25; 31 female and 34 male). All participants were either born in Algiers or came to Algiers at an early age. They all had Algerian Arabic as their mother tongue and either started learning Standard Arabic when they started school or at kindergarten or mosque. However, participants differed in the time of acquiring French. Early bilinguals reported that they started French together with Algerian Arabic or shortly after.

They also said that they watched cartoons mostly in French. Late bilinguals started French at school at around age 8, and may or may not have watched cartoons in French before they started French at school. All Bilinguals also claimed that they code-switch with friends and family, but they differ in the time when they started code-switching or in how often they code-switch. Participants were recruited by means of an announcement for the study via leaflets containing conditions for participation and were paid for their participation.

### Language Proficiency Assessment

To assess French proficiency, participants completed the French Cloze test developed by Tremblay (2011). The test consists of a text that contains blanks (deleted words) and which the participants had to read and fill in each blank with one word. Of the 45 blanks, 23 were content words (e.g., nouns, main verbs, adjectives, etc.) and 22 were function words (determiners, pronouns, prepositions, etc.). Standard Arabic proficiency was assessed using a Cloze test developed for the purpose of this study. The test contains 35 deleted words of which 26 were content words and 9 were function words. In order to standardize the Arabic Cloze test, 12 Algerian speakers who did not take part in the actual study, completed the test. From their responses, a bank of acceptable answers was created and used for scoring the test. The scores from both tests were converted into percent accuracy rates.

### Language History and Switching Habits Questionnaire

The participants completed a language questionnaire. This was a French translated version of “*The assessment of code-switching experience survey*” (ACSES) developed by Blackburn and Wicha (2011). The questionnaire starts with some autobiographical questions concerning the participants' age, gender, place of birth, and residence. The participants also self-rated their proficiency in reading, writing, speaking, and listening for Algerian Arabic, Standard Arabic, and French. The questionnaire included multiple choice questions concerning the participants' daily use of languages, their code-switching habits, reasons for code-switching and their attitudes toward code-switching. Code-switching scores are the averages of daily use of languages and frequency of code-switching.

In addition to these tests participants were given a semantic fluency test, the Simon task, a working memory test, and an interview. These data will not be reported here.

### Material and Design

The stimuli contained a total of 32 non-cognate French target words (underscored in **Table 1**). We used non-cognate nouns because it would be hard to find cognate nouns that are shared between Algerian Arabic, Standard Arabic, and French. Cognates are typically loans from French into Algerian Arabic but not into Standard Arabic. The French words were embedded in high and low-constraint Algerian Arabic and Standard Arabic sentences. Stimuli, thus, included 16 AA-FR code-switched sentences and 16 SA-FR code-switched sentences. The cloze probability of the sentences was determined on the basis of a web-based completion

TABLE 1 | Sample of experimental item set.

Condition	Sample sentence
AA base language High-constraint (AAH)	Kul-ma naʔaslu ssaan lazem nʔallu <b>la bouche</b> . “Every time we wash the teeth we rinse <u>the mouth</u> .”
SA base language High-constraint (SAH)	fi kuli maratin naʔsilu fiha ʔal ʔasnaa jadʔibu ʔan naʔtʔifa <b>la bouche</b> . “Every time we wash the teeth we rinse <u>the mouth</u> .”
AA base language Low-constraint (AAL)	Had l'ewled ma rʔadsh ʔlaxatʔer ʔ kan ʔendu sʔtʔar fi <b>la bouche</b> . “This boy did not sleep because he had pain in <u>the mouth</u> .”
SA base language Low-constraint (SAL)	ʔina haaʔ ʔal walad lam janam liʔanahu kaana juʔaani min ʔalamin fi <b>la bouche</b> . “This boy did not sleep because he had pain in <u>the mouth</u> .”

study on 76 Algerian bilinguals not participating in the actual study. The sentences were initially presented in French leaving the final noun phrase out for the participants to complete with three possible best completions. The mean cloze probabilities for the target words in the high-constraint sentences was (0.77), and in the low-constraint sentences was (0.06). Sentences in Algerian Arabic and those in Standard Arabic were close translations to the French sentences as reviewed by three Algerian bilingual speakers. In all sentences, the target words were French noun phrases formed with a feminine noun and a feminine definite article that appeared at the end of the sentence. Masculine nouns were avoided because the French masculine article *le* preceding the noun is mostly replaced by the Algerian Arabic article *əl* which needs to be assimilated to the initial consonant of the noun when it is a *solar*, i.e., a coronal consonant. Noun phrases in this study were visually presented in French. Thus, in the case when the context constrains toward a masculine noun that starts with a coronal sound, participants may anticipate the assimilated article *əl* but find the unassimilated article *le* instead. This might incur extra processing that confounds with the processing that is due to language switch expectation in the controlled conditions. The mean length of sentences was (mean = 8, range = 6–11 words) and in milliseconds (mean = 3071 ms, range = 1756–4705 ms). Experimental target nouns were controlled for length and frequency for the purpose of Latin Square (see below): length (mean = 8, range = 6–11 characters) and frequency (mean = 6, range = 4–8). The frequency of the French nouns was based on an online survey completed by twelve Algerian bilinguals. A sample sentence in the four conditions is displayed in **Table 1**.

Another 64 sentences were constructed as fillers. Half of the fillers used Algerian Arabic and half used Standard Arabic. To avoid the adoption of a strategy by the participants, half of the fillers had switches at different points of the sentences. Filler switches appeared either earlier toward the beginning of the sentences, in the middle or toward the end of the sentences but never word finally. The other half did not contain switches and were, therefore, only heard. The filler switches were either nouns, verbs, adjectives or adverbs. All experimental and filler sentences contained switching from Arabic to French because this type is more common than switching from French to Arabic. In addition, Algerian Arabic is not traditionally written, rendering Algerian Arabic language unsuitable for the targets in the present paradigm. Four lists of stimuli were constructed using Latin Square, such that each list contained one sentence in each of the four conditions, and no list contained more than one version of

each sentence. Each participant saw only one list of 96 sentences and each experimental target NP appeared only once in each list. The fillers were the same across the four lists. Sentences in each list were pseudo-randomized to avoid order effect, and lists were randomly assigned to participants. All sentences were recorded by the same bilingual Algerian female speaker in a soundproof boot using a Marantz PMD660 digital recorder, recording 16-bit stereo PCM sound at a sampling rate of 44.1 kHz. In order to minimize co-articulation, a dummy word “*huda*” was inserted instead of the actual target words. The auditory sentences were coded and segmented using (Boersma, 2001): the speech signal was cut just before the presentation of the target word, using nearest zero crossing selection, and the durations of the speech segments were extracted. The auditory sentences were then normalized to minimize the difference in amplitude across all sentences.

## Procedure

Once arrived to the study site, the participants first completed the French proficiency cloze test, then completed the following tests not included in the analysis of the current study: a French semantic fluency test, the Simon task, the memory test and the interview in this order. After a short microphone test, participants started the CMN experiment with a practice session. The practice task consisted of five sample trials resembling the experimental and filler sentences, that is, the target words appeared at the end of a sentence, somewhere in the middle or the trial did not have any visual target words. During the practice session the experimenter remained next to the participants and gave feedback on their performance. When the participants started the experiment, the experimenter remained in the room but withdrew to a corner. After the completion of the naming experiment, the participants completed the cloze test for Standard Arabic proficiency followed by the Arabic semantic fluency task (not included in this analysis), and the language history and switching habits questionnaire.

Stimuli in the CMN task were presented on the screen of a laptop computer using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, USA). Participants were seated at about 50 cm from the computer with a microphone on a stand sitting in between, and a response button box on the right side of the computer. Participants also wore a headphone set with a microphone attached to the computer digital array mic. While the headphone presented the auditory stimuli, the head-mounted microphone recorded the participants' naming

responses. Reaction times to targets naming were collected using a voice trigger via the standing microphone attached to the response button box. The participants were instructed to listen carefully to the content and read the words that appeared on the screen as quickly and accurately as possible. To make sure the participants paid attention to the content, they were told they would be asked some questions at the end. Participants had to press any button on the response box to proceed to a trial. First, the participant saw a *Prêt* “Ready” sign at the center of the screen and pressed a button to proceed when they were ready to start. The ready sign then changed to a cross sign “+” inside a rectangle and remained on the screen while the participants listened to the auditory part of the sentence. At the offset of the last word, the cross sign disappeared and the target word was displayed instead for a duration that was equal to the word length (in milliseconds) times 150 plus 1200 ms. The participants were asked to read the words out loud. When the display time ended the “Ready” sign started the following trial in the case of an experimental trial. In the case of a filler trial, in which the naming occurred somewhere in the middle of the sentence, the cross sign was displayed again and the participants heard the final auditory part that completed the sentence before they saw the “Ready” sign again. When the fillers did not contain a target word to read, participants heard the entire sentence until they saw the “Ready” sign to proceed to the next trial.

## RESULTS

The analysis was conducted on correct responses only. A correct response is one which was clearly fluent with no stammering or hesitation. Correct responses after self-correction were not accepted. Answers which were ambiguous due to unclear pronunciation or lack of audibility were presented to another Algerian-French bilingual speaker. If the bilingual speaker could not identify the word or hesitated about identifying it, that word was excluded from the analysis. Raw data consisted of 2080 data points. Incorrect or non-identifiable responses constituted 1.3% (27 data points). Accurate responses were equally distributed across the four conditions: Algerian Arabic high-constraint: AAH (98%); Standard Arabic high-constraint: SAH (99%), Algerian Arabic low-constraint: AAL (99%); Standard Arabic low-constraint: SAL (98%). Overall accurate data was 2053 data points (98.7%). We conducted all analyses on the log transformed residual reaction times in order to control for potential effects of target word length, frequency and the position of the trial in the experiment. Log transformed reaction times were residualized for length (in characters) and frequency of the target words, and for the trial number, that is, the order in which a certain item appeared in the experiment. Residuals were calculated by means of a linear mixed effect model conducted on the log transformed response times in the experimental trials, with target word length (in number of characters), target word frequency, and trial order as fixed effects, and a by-subject random intercept. Response times estimated by this model were then subtracted from the log transformed response times to obtain the log residual response times. Outliers were then removed from each condition for each

participant using the mean  $\pm$  3 standard deviation method. After cutting off for outliers, 2028 data points (98.8%) remained that is 25 data points (1.2%) were removed.

In order to determine which factors to include in the model, tests of correlations using the `rcorr()` function in the `Hmisc` package were utilized in order to further explore the correlations/covariances and significance levels for Pearson and Spearman correlations between code-switching habits (as measured by the ACSES questionnaire), age of acquisition of French and proficiency in French (as determined by the Cloze test). There was a medium-sized negative correlation between Age of acquisition and proficiency,  $r = -0.36$ ;  $p < 0.01$ . The later French was acquired, the less proficient the bilingual. However, correlations between code-switching and age of acquisition, and code-switching and proficiency did not reach significance: with age of acquisition  $r = 0.09$ ;  $p = 0.49$ ; with proficiency,  $r = -0.09$ ;  $p = 0.47$ . Based on these correlation results, proficiency but not age of acquisition was included as a factor in the analysis of the target word naming times.

Naming latencies to the French target words were then analyzed using a linear mixed effects model `lmer` in R (version 3.1.3, R Core Team, 2015) as implemented in the package `lme4` (version 1.1-7, Bates et al., 2015). The model included semantic constraint (High Cloze/Low Cloze, with “high cloze” coded as  $-0.5$  and “low cloze” as  $0.5$ ), base language (AA/SA, with “AA” coded as  $-0.5$  and “SA” as  $0.5$ ), and the continuous variables French Proficiency (`FrProf`) and code-switching habits (CS) as fixed effects. In addition, the interactions between each two of these factors (except that between proficiency and code switching) were also included as fixed effects. The analysis contained a random effect structure which included by-subject and by-item random intercepts, with the fixed effects constraint, base language and their interactions as by-subject and by-item random slopes. The fixed effects were group-mean centered to minimize collinearity. After centering, the maximal variance inflation factor was 1.04, and there were no signs of collinearity in the analysis (fixed effect correlations  $r_s < 0.2$ ). Significance of the fixed effects was obtained on the basis of the  $t$ -values of the estimates of the coefficients. Absolute  $t$ -values of 1.96 or larger were considered significant. To explore significant interactions produced by the model, we followed-up with separate linear mixed-effects models fitted for each specific group. In addition, we calculated effect sizes “Cohen’s  $d$ ” following Dunlap et al. (1996) to compare groups on the basis of the magnitude of a statistically significant effect and we provided 95% confidence interval (Table 5). Note that the analysis was conducted on the log transformed residualized reaction times, however, the means in the main text are given for the raw reaction times for the reader’s convenience.

## Overall Analysis

A maximal model was fitted and the model converged without simplifying the random slope structure. Analysis on the residual RTs for the entire group of participants as summarized in Table 2 shows a main effect for semantic constraint, a semantic constraint by code-switching habits interaction, and a semantic constraint by proficiency interaction. However, base language effect was not

**TABLE 2 | Results of the residual naming latencies mixed effects analysis for whole group.**

Effect	$\beta$	SE	t-value
Intercept (mean)	-0.006	0.007	-0.83
Semantic constraint	0.03	0.01	2.78*
Base language	0.01	0.01	1.02
French proficiency	0.0005	0.0005	1.13
Code-switching	0.004	0.005	0.89
Semantic constraint*Base language	-0.02	0.02	-0.93
Semantic constraint*Code-switching	-0.02	0.01	-1.98*
Semantic constraint*French proficiency	-0.002	0.001	-2.64*
Base language*Code-switching	-0.01	0.01	-1.64
Base language*French proficiency	0.0003	0.001	0.28

SE, standard error; \*significant t-value ( $p < 0.05$ ).

significant and no interaction with base language was significant. We will elaborate on each of these effects below in the light of the research questions.

### Language Switch Expectancy and Semantic Constraints

There was a significant main effect of semantic constraint (Table 2). Participants' naming latencies were shorter in the high-constraint sentences ( $M = 565$  ms,  $SD = 158$ ) than in the low-constraint sentences ( $M = 583$  ms,  $SD = 179$ ), suggesting that it was easier for the participants to anticipate the target words more when the preceding context provided rich semantic clues about the forthcoming lexical items.

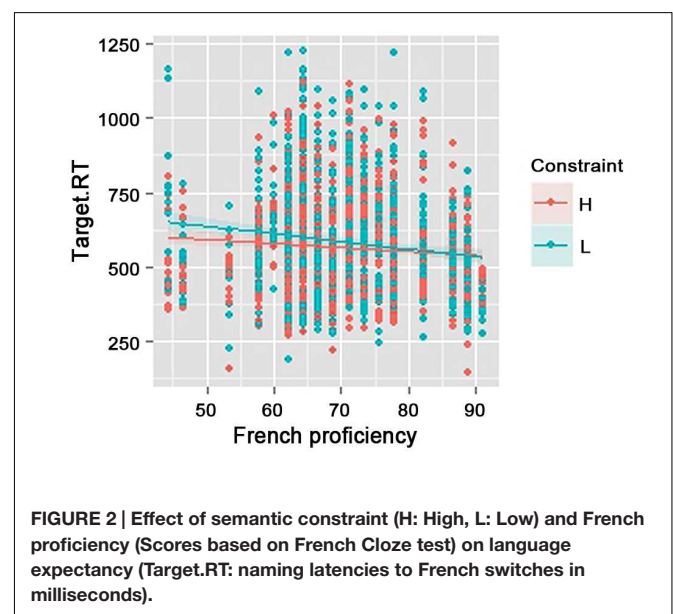
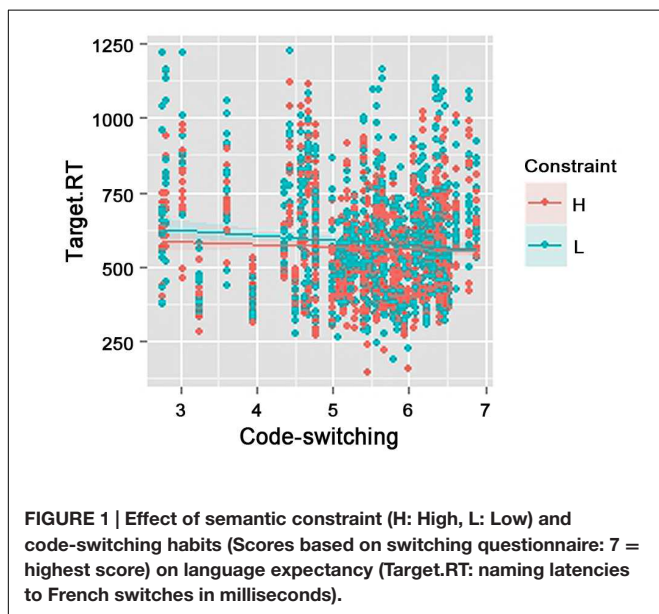
The interaction between semantic constraint and code-switching habits was significant (Table 2). As shown in Figure 1, the semantic effect (faster naming latencies for high-constraint than low-constraint sentences) was larger for the participants who code-switched less frequently than for those who code-switched more frequently.

Semantic constraint also interacted with French proficiency (Table 2) revealing a significant difference between high proficiency bilinguals and low proficiency bilinguals in the effect of semantic constraint of a sentence context. Overall, highly proficient participants were faster and showed reduced semantic constraint effect than low proficient participants (Figure 2), suggesting that bilingual code-switchers who are more proficient in L2 French tend to expect switching into French more than code-switchers of lower L2 proficiency.

### Exploring the Interactions Semantic Constraints and Code-Switching Habits

In order to further explore semantic constraint by code-switching habits interaction, we examined the bilinguals at the two extremities of the continuous code-switching line. We compared bilinguals with the lowest code-switching scores to bilinguals with the highest code-switching scores in the overall group. Twenty participants in each code-switching group type (light code-switchers/heavy code-switchers) were selected based on their code-switching scores in the language history and switching habits questionnaire (ACSES). The code-switching score was calculated based on the averages between daily use of languages and code-switching habits. The code-switching groups were matched on age, language proficiency and age of acquisition. Table 3 reports the characteristics of the participants in the code-switching groups.

Treating code-switching as categorical, we constructed a linear mixed effect lmer for the heavy code-switchers and a separate lmer for the light code-switchers. Both lmer models contained constraint (High Cloze/Low Cloze, with "high cloze" coded as -0.5 and "low cloze" as 0.5), base language (AA/SA, with "AA" coded as -0.5 and "SA" as 0.5), the continuous variable French Proficiency (FrProf) and the interactions between each two of these factors as fixed effects. The random effects structure was





**TABLE 3 | Participant characteristics in the code-switching groups.**

Characteristics	Code-switching group		p-value
	Light code-switchers CS ≤ 5.18, n = 20	Heavy code-switchers CS ≥ 6, n = 20	
Code-switching	4.29	6.37	0.000*
Age	22.05	22.00	0.90
French proficiency	73.44	71.44	0.53
Age of FR acquisition	6.30	7.15	0.21
Standard Arabic proficiency	74.55	76.76	0.54
Age of SA acquisition	5.10	5.15	0.81

\*Significant p-value < 0.05.

similar to that in the overall analysis model. As before, the fixed effects were centered to minimize collinearity. After centering, the maximal variance inflation factor was smaller than 1.05, and there were no signs of collinearity in the analysis (fixed effect correlations  $r_s < 0.2$ ).

The analysis from both lmer models for the code-switching groups separately showed a significant main effect of semantic constraint in the light code-switching group: [ $\beta$ : 0.04, SE: 0.01,  $t$ : 2.48], but not in the heavy code-switching group [ $\beta$ : 0.01, SE: 0.02,  $t$ : 0.81]. Comparison of the effect sizes confirmed that the light-code switchers showed a larger effect of semantic constraint (Cohen’s  $d$ : 0.59) than the heavy code-switchers (0.17). The difference between high and low constraints was larger for the light code-switchers (naming took 24 ms longer in the low than the high constrained context) than for the heavy code-switchers (10 ms). In particular, the means of naming latencies in high and low-constraint contexts revealed that light code-switchers differed from heavy code-switchers in the highly constraining context. Light code-switchers had shorter naming latencies ( $M = 576$  ms,  $SD = 184$ ) than heavy code-switchers ( $M = 595$  ms,  $SD = 149$ ) indicating that they processed the switch faster. In the low constraining context, naming latencies in light code-switchers ( $M = 600$  ms,  $SD = 209$ ) were not different from those in heavy code-switchers ( $M = 605$  ms,  $SD = 170$ ). This may also suggest that bilinguals in the light code-switching group anticipated a language switch compared to bilinguals in the heavy code-switching group.

**Semantic Constraints and French Proficiency**

The interaction between semantic constraint and proficiency was explored by comparing the effect of semantic constraint in the lowest proficient bilinguals and the highest proficient bilinguals from the overall group. Based on scores in the French proficiency test, two groups (low proficient/high proficient) were selected, each containing 20 participants. The high proficiency group started learning French at an earlier age than the low proficiency group (Table 4), but the two proficiency groups were matched on age, code-switching habits, proficiency in Standard Arabic and age of Acquisition of Arabic.

We constructed two linear mixed effects lmer models separately for the two proficiency groups with proficiency treated as categorical to examine semantic effect significance. The models contained constraint (High Cloze/Low Cloze, with “high cloze” coded as  $-0.5$  and “low cloze” as  $0.5$ ), base language (AA/SA, with “AA” coded as  $-0.5$  and “SA” as  $0.5$ ), code-switching habits (CS) as a continuous variable, and the interactions between each two of these factors as fixed effects. The random effects structure was similar to that in the overall analysis model. The fixed effects were centered to minimize collinearity. The maximal variance inflation factor after centering was smaller than 1.04, and there were no signs of collinearity in the analysis (fixed effect correlations  $r_s < 0.2$ ).

The effect of semantic constraint was still significant in the low proficiency group: [ $\beta$ : 0.06, SE: 0.02,  $t$ : 2.92]. Naming latencies in the low proficient bilinguals were 32 ms shorter in the high-constraint context ( $M = 571$ ,  $SD = 172$ ) compared to the low-constraint context ( $M = 603$  ms,  $SD = 208$ ). However, the effect of semantic constraint in the high proficiency group was not statistically significant: [ $\beta$ : 0.02, SE: 0.02,  $t$ : 1.08]. High proficiency bilinguals were only 10 ms faster in responding to the targets in the high ( $M = 523$ ,  $SD = 135$ ) than in the low ( $M = 533$ ,  $SD = 145$ ) constraint contexts. Effect sizes confirmed these results: there was a larger effect size in the low proficiency group (0.7) but a relatively small effect size in the high proficiency group (0.31).

**Effects of Base Language**

Although the analysis does not show an interaction between semantic constraint and base language we wanted to explore the effect of semantic constraint in each base language separately given that we hypothesized that semantic constraint effect should be more visible in Algerian Arabic because of the high expectation of a French continuation in daily language use. Separate lmer models for Algerian Arabic base language trials and Standard Arabic base language trials revealed a significant main effect of semantic constraint in Algerian Arabic base language: [ $\beta$ : 0.04, SE: 0.01,  $t$ : 2.53], but not in the Standard Arabic base

**TABLE 4 | Participant characteristics in the proficiency groups.**

Characteristics	Proficiency group		p-value
	Low proficient Accuracy < 65%, n = 20	High proficient Accuracy ≥ 77%, n = 20	
Code-switching score	5.36	5.24	0.72
Age	22.05	21.60	0.34
French proficiency	60.11	82.89	0.000*
Age of FR acquisition	7.50	5.40	0.001*
Standard Arabic proficiency	73.38	75.44	0.56
Age of SA acquisition	4.85	5.05	0.40

\*Significant difference between the two groups (p-value < 0.05).

**TABLE 5 | Ninety five percentage confidence interval of the mean differences in raw RTs for semantic effect in overall and group analyses.**

Semantic effect by group	95% confidence interval		Semantic effect by base language	95% confidence interval	
	2.5	97.5		2.5	97.5
Overall	-32.79	-3.32	Overall AA	-47.57	-10.62
Light code-switching	-55.77	-6.88	Overall SA	-29.39	4.37
Heavy code-switching	-37.52	15.33	Light code-switching AA	-80.58	-4.97
Low proficiency	-71.53	-11.73	Light code-switching SA	-52.59	13.35
High proficiency	-24.98	5.13	Low proficiency AA	-104.40	-4.143
			Low proficiency SA	-68.96	16.96

language: [ $\beta$ : 0.02,  $SE$ : 0.01,  $t$ : 1.635]. Naming latencies to the French targets were shorter in the high semantically constraining sentences ( $M = 558$  ms,  $SD = 150$ ) than in the low semantically constraining sentences ( $M = 584$  ms,  $SD = 184$ ) when the base language was Algerian Arabic. However, when the base language was Standard Arabic naming latencies in the high constraint context were not significantly faster ( $M = 571$  ms,  $SD = 166$ ) than those in the low constraint context ( $M = 581$  ms,  $SD = 174$ ). The results suggest that participants benefited more from semantic manipulation by using the semantic cues in the high-constraint sentences during listening to Algerian Arabic compared to listening to Standard Arabic. The effect sizes confirmed this interpretation. Although effect sizes are both relatively small Cohen's  $d$  was smaller in the Standard Arabic base language (0.18) than it was in Algerian Arabic base language trials (0.36).

The effect of semantic constraint in the code-switching groups was found to be larger for the light code-switchers. We conducted separate analyses by base language in order to see in which base language was the effect size more important. Comparison of the analyses for Algerian Arabic base language trials and Standard Arabic base language trials in light code-switchers showed a significant main effect of semantic constraint in Algerian Arabic base language: [ $\beta$ : 0.05,  $SE$ : 0.02,  $t$ : 2.28], mean naming latencies in the high-constraint context ( $M = 562$ ;  $SD = 163$ ), and in the low-constraint context ( $M = 596$ ;  $SD = 209$ ), but not in Standard Arabic base language: [ $\beta$ : 0.03,  $SE$ : 0.02,  $t$ : 1.13], mean naming latencies in the high-constraint context ( $M = 591$ ;  $SD = 202$ ), and in the low-constraint context ( $M = 604$ ;  $SD = 211$ ). Comparison of the effect sizes confirmed these results: Cohen's  $d$  was (0.46) in Algerian Arabic base language but it was (0.27) in Standard Arabic base language. Once again, the results suggest that the effect of semantic constraint was driven by the context language that is commonly used in interactional contexts and is part of the more typical AA-FR code-switching.

Similarly, the effect of semantic constraint was larger in the low proficiency group. Analysis by base language in the low proficiency group revealed a significant semantic constraint effect in Algerian Arabic base language trials: [ $\beta$ : 0.08,  $SE$ : 0.03,  $t$ : 2.25], mean naming latencies in the high-constraint context ( $M = 564$ ;  $SD = 163$ ) were 45 ms shorter than those in the low-constraint context ( $M = 609$ ;  $SD = 219$ ). However, semantic constraint effect was not significant in Standard Arabic base language trials: [ $\beta$ : 0.04,  $SE$ : 0.03,  $t$ : 1.21], mean naming latencies in the high-constraint context ( $M = 578$ ;  $SD = 181$ ) were 19 ms shorter than naming latencies in the low-constraint context ( $M = 597$ ;

$SD = 196$ ). Comparison of the effect sizes showed a larger effect size in Algerian Arabic base language (0.48) than in Standard Arabic base language (0.26).

These results are in line with our prediction that the difference in semantic constraint effect should be seen more in the typical AA-FR code-switching context than in the atypical SA-FR code-switching context. The fact that naming latencies were constantly shorter and the effect sizes constantly larger in Algerian Arabic than in Standard Arabic is evidence that base language did affect the processing of the switch. Semantic facilitation in the high-constraint context in Algerian Arabic base language promoted the processing of a code-switch. In particular, the results suggest that when the switch is part of the typical language pair that is repeatedly used in conversation, its processing is easier. It may also suggest that participants could anticipate a language switch when they heard Algerian Arabic. An observation worthy of notice is that Algerian Arabic and standard Arabic are rather similar in several aspects, and thus differences in effect sizes should not be expected. In this respect, the observed differences between Algerian Arabic and standard Arabic base languages in the different groups, though not always large, are informative for models of code switching.

## DISCUSSION

In this study, we sought to examine the effect of language use and semantic constraints on the expectancy of a language switch during listening comprehension in Algerian bilingual speakers. In particular, expectation of a language switch was compared between two types of code-switched sentences that involved different pairs of languages/varieties. The first occurs between Algerian Arabic and French and is typically conversational and frequent among the bilingual community that code-switches. The second type involves code-switching between Standard Arabic and French which is neither interactional nor typical of Algerian bilinguals. Participants heard the first part of the code-switched sentences presented either in Algerian Arabic or Standard Arabic then, immediately after, read a French NP that completed the first parts. Naming latencies to the French NPs were measured and compared. Faster reaction times suggested an easier processing of the target word, which can be interpreted as a higher expectation of a language switch and ease of switch processing. We asked (1) whether language expectancy in code-switching depends on the base language; (2) whether semantic constraints affect language

expectancy in code-switching; (3) whether language expectancy is dependent on the frequency of code-switching; and (4) whether French L2 proficiency modulates the expectancy of language switching.

The findings revealed three effects: semantic constraint effect; an interaction between constraint and code-switching habits; and an interaction between constraint and French proficiency. Bilinguals were significantly faster in the high than in the low-constraint context, suggesting that a language switch is more expected and/or the switch is easier to process when it is supported by the semantic information of the sentence context. This also suggests that the CMN task was sensitive to sentence context and to lexical activation. In addition, the semantic constraint effect, that is, the difference between reaction times in high and low-constraint contexts, was larger when the base language was Algerian Arabic than when it was Standard Arabic. This suggests that the listeners made more use of the semantic cues provided by the high-constraint context in the more typical code-switching that is more recurrent in the everyday interactions. In addition, the frequency of daily code switching modulated the effect of semantic constraint of a sentence context. Light code-switchers but not heavy code-switchers were significantly faster in the high-constraint context than in the low-constraint context preceding the switch. This suggests that the habit of switching between languages interferes with our predictions and with the state of activation of both languages. However, these results look counterintuitive. One would assume that the more a bilingual code-switches the more he/she expects a language switch. We will provide a speculative interpretation of this below. Finally, we found that high proficiency bilinguals had shorter naming latencies than low proficiency bilinguals. French proficiency modulated the effect of semantic constraint on language switch expectancy. Bilinguals with low proficiency in French showed larger constraint effect, with faster reaction times in high-compared to low-constraint context. As proficiency increased the difference in naming latencies between high and low-constraint contexts became smaller, probably due to the overall increase in speed, leading to a reduced effect of sentence context.

## Theoretical Accounts and Implications

The major finding of the current study is that the effect of semantic context is contingent on the bilingual's language use. In particular, the effect of semantic context occurred to the extent to which the bilinguals code-switch in everyday interactions. Semantic constraint effects were reduced in bilinguals who frequently code-switch, but were visible in bilinguals who code-switch less frequently. Studies reporting reduced sentence influence in the high-constraint context (e.g., Altarriba et al., 1996; Titone et al., 2011) suggested that the readers in the highly constrained context generate semantic as well as lexical features of the upcoming words in the context language. However, when the target words mismatch the expected words in phonology, there is extra processing. They also suggested that the semantic context can selectively activate a word in one language. Titone et al. (2011) noted that L1 can activate L2 in a highly constraining context when L2 phonology is salient (e.g., when sentences from

L1 and L2 are intermixed in the study), and when the bilinguals are highly proficient in their L2. In a recent study, Boukadi et al. (2015) tested Tunisian Arabic- French bilinguals who are moderately proficient in French. The study used picture-word interference task in monolingual and bilingual contexts. When the context was monolingual, naming latencies were affected only by phonological facilitation, suggesting that lexical selection proceeded in a selective manner. In the bilingual context, a phono-translation effect as well as phonological and semantic effects were found suggesting that lexical selection is non-specific. The writers suggest that the bilingual lexical selection is dynamic and depends on factors such as the experimental language context (monolingual or bilingual). The degree of activation of both languages determines whether lexical selection functions in a language non-specific or in a language-specific manner. If the extra processing in the heavy code-switching group is due to the fact that the participants generated lexical features of the upcoming words in the context language, then the question is why participants in the light code-switching group did not generate lexical features in the context language. If on the other hand, facilitation in the high-constraint context in the light code-switching group is attributed to the activation of L2, then the question is why the heavy code-switchers did not activate L2 to the same extent as the light code-switchers did. Since participants in the two groups differed only in the frequency of daily code-switching, the different results may be related to the habit of switching between languages.

One of the bilingual language processing models that account for sentence context influence is the Bilingual Interactive Activation+ (BIA+; Dijkstra and Van Heuven, 2002). The BIA+ assumes that words from both of the bilingual's languages are integrated in one lexicon in which activation is parallel and language non-selective. The fact that sentence context influenced reaction times to the French target words differently in low and high-constraint contexts is itself in support of the BIA+ assumptions regarding the influence of a sentence context on word recognition in bilinguals. The model assumes that sentence context affects word recognition through increased activation of items semantically related to the context. Because lexical activation is non-selective, all words that meet the semantic features evoked by the high-constraint context are activated regardless of their language membership. The activated words should then compete for selection. Competition is resolved by the top-down decision system that inhibits the task schema of the activated words in the non-intended language. In the case of the current study, words in Algerian Arabic and Standard Arabic should be inhibited in order for the bilingual to be able to name the words in the intended French language. This process should incur extra processing in code-switching the more strongly the anticipated word candidates are activated in the context language. Assuming that bilinguals build strong predictions in the highly constraining context, then competition between the activated words in the base languages and the words to be named in French would be strong and lexical selection through inhibition would incur extra processing load, leading to longer naming latencies. In this case, constraint effect is predicted to become smaller as naming latencies in the high-constraint

context become larger. This is not supported by the present data which show reduced semantic constraint effects in the heavy code-switching group, but large effects in the light code-switching group due to facilitation effect in the high-constraint context. The finding that sentence context effects were different for the code switching groups is therefore somewhat problematic for this view.

The BIA+ model also contains a layer of two language nodes that function as language tags showing the membership of a word. The language nodes become activated late in the process and do not directly influence the lexical candidates. The model recognizes that the presence of a sentence context can pre-activate the language nodes, but because the language nodes cannot inhibit the non-target language words completely, sentence context cannot restrain language non-selective activation. The model does not clearly indicate the mechanism by which sentence effect takes place. With the assumption that the language nodes are activated late and cannot directly influence the lexical candidates, boosted semantic activation by itself is not enough to explain the different effect of sentence context in both groups. To account for the absence of cognate facilitation in the high but not in the low-constraint context, Schwartz and Kroll (2006) suggests that language nodes can be pre-activated by the highly constraining context. This may occur when the increasing constraining context attains an early stable activation in the lexicon which allows for an earlier activation of the language nodes. Even an additional assumption of the pre-activation of the language nodes through the increasing contextual constraint will not account for the different results in both code-switching groups. In that case, we should be looking for an explanation for why heavy code-switchers pre-activated the language nodes but not light code-switchers. In the rest of this paper, we will discuss how the findings in this study may be accounted for by a recent model of code-switching (Green and Wei, 2014).

The central idea of the control process model of code-switching is that language control varies depending on the different interactional contexts of the bilingual speaker and that the processes of language control can adapt to the demands imposed on them by these different interactions. The findings in the current study reveal that the effect of semantic constraint on the naming latencies to the French switches depended on the frequency of daily code-switching. Heavy code-switchers were slower to name the French NPs in both constraint conditions. In addition, the size of constraint effect was very small in bilinguals who frequently code-switch but was larger in bilinguals who do not code-switch frequently. This finding supports the general assumption of the control process model of code-switching that different language contexts induce different habits of language control. Shorter naming latencies in the high-constraint context may suggest that light code-switchers expected more a switch to the other language or that they integrated the switch more easily than heavy code-switchers. However, this interpretation sounds counterintuitive. Bilinguals who frequently code-switch should be more prepared to hear or integrate a switch. How can these results be interpreted by the control model?

The adaptive control hypothesis makes two important assumptions. The first states that experimental contexts can

trigger the types of control processes in bilinguals. The prediction that follows is that the bilingual's reactions to an experimental context can vary depending on how well the context fits the type of control processes for that bilingual. The second assumption concerns the individual differences. While bilingual speakers may experience more than one type of interactional contexts, their dominant type of language control is contingent on the typical exchanges that are recurrent within their speech of community. The model thus predicts that bilinguals who experience different interactional contexts may show adaptive responses that vary depending on how typical they are of each interactional context. The type of bilinguals tested in the current study are more representative of the dense code-switching in the Green and Wei model because they tend to adapt words morphologically as well as phonologically in informal contexts, although they may use French only during classroom hours, or use insertion in some other contexts. However, some of those participants are better representatives of dense code-switchers than others. The light code-switchers use both languages daily but do not frequently code-switch; their switches may be more regarded as insertions rather than integrated switches. The bilinguals who claimed they code-switch regularly may include more integrated switches than bilinguals who code-switch less frequently. In fact, some bilinguals asked during the training session whether they should read the target words in Algerian Arabic (meaning with Arabic phonology) or in French. Interestingly, even reminding the participants to read the targets in French did not eliminate few errors of the kind of phonological integration. In this case, those who frequently code-switch may be more familiarized with control processes that permit opportunistic planning, but those who do not code-switch frequently, and yet use both languages daily, may be more trained with interference suppression that taps on competitive relationship between the language schemas. On the other hand, the stimuli in the current study include code-switches that are in the form of insertion, baring no syntactic or phonological adaptation in the base language structure. When these stimuli are encountered, the language schemas are forced to consistently restrain from adapting the words. The stimuli may also require the participants to be in a coupled mode in which control passes from one schema to the other. In this case, the bilinguals who do not frequently code-switch may be more used to the type of stimuli presented in this study. The results revealed that light code-switchers showed larger semantic constraint effects that reflects facilitation of response in the high compared to the low context. This may suggest that the stimuli triggered the control processes which the light code-switchers practiced more in their daily interactional contexts. By contrast, heavy code-switchers took longer time to name the switches and showed reduced semantic effects suggesting that facilitation in the high-constraint context did not occur. The stimuli should have forced them to engage control processes that are not typical of their interactional context. Heavy code-switchers had to control for the target words adaptation and engage a coupled control needed for insertion, whereas they are more used to an open control in which adaptation is allowed.

We turn now to consider the effect of base language. A main question in this study was to determine the effect of the languages

involved in code-switching on the expectancy of a language switch. Comparison of effect sizes showed systematically a larger effect of sentence constraint in Algerian Arabic base language than in Standard Arabic base language. The results suggest greater expectancy of a switch when it is part of the typically conversational, recurrent code-switching. These results are even more important taking into account the relationship between Algerian Arabic and Standard Arabic. Out of 32 sentences, tested in this study, that occurred in the high constraining context 22 sentences were biased toward a target continuation that is shared between Algerian Arabic and Standard Arabic, assuming that the items are indeed predicted in the same language of the context. For instance, if a sentence context in Algerian Arabic constrained toward the Algerian Arabic *el baab* “the door,” a similar sentence context in Standard Arabic would bias toward the Standard Arabic *el baab* “the door.” It should be predicted that the participants’ reactions to the French translation NP *la porte* “the door” would be the same when listening to that sentence context in Algerian Arabic or in Standard Arabic. Given this information, one would not predict a difference between the two types of code-switching had the bilinguals anticipated the continuations in the base languages. In this respect, the differences in the naming latencies found between targets in the Algerian Arabic and Standard Arabic trials are to some extent meaningful. In particular, they suggest that it is possible for these bilinguals to anticipate a switch to L2 French when they listen to an Algerian Arabic context than when they listen to a Standard Arabic context. However, these results may be interpreted as ease of integration of a French switch when this later is part of the conversational code-switching.

The results concerning base language effect may be better understood by considering how the three languages are connected and stored in the bilingual memory. Both code-switching groups had moderately advanced level of proficiency in French and Standard Arabic. They acquired French at about 6/7 years of age and Standard Arabic at about 5. However, the groups differed in the frequency of code-switching. The findings suggest that in the light code-switchers, the French target item was already activated in the high-constraint context with the other translations in the base languages. When the bilinguals named the words in French they benefited from the early activation leading to a faster lexical retrieval. For the heavy code-switchers, activation of the French forms may not be as simultaneous as the other languages. Thus, naming the French targets would require extra time. The organization of lexical items may differ greatly depending on how the bilinguals represent the words semantically across the languages they speak (e.g., Basnight-Brown, 2014). There is a long line of literature showing that the ease of lexical retrieval depends on the degree of proficiency in L2 (e.g., Heredia, 1997; Heredia and Altarriba, 2001; Sunderman and Kroll, 2006; Kroll et al., 2010). When bilinguals become more proficient in L2, dominance may shift in some areas and words become more accessible in the language the bilinguals most often use (e.g., Cieślicka and Heredia, 2015). Results from the proficiency groups support

this idea. Lexical retrieval in high proficient bilinguals was much faster than in low proficient bilinguals regardless of the context constraint, suggesting that activation is more automatic. However, proficiency by itself cannot account for differences among the code-switching groups in which bilinguals have about the same degree of proficiency, regardless of their code-switching habits. Language usage, and in particular the manner and the frequency of switching between languages, may have affected their lexical organization across these languages and hence lexical retrieval. Contrasting the present results with those from a study that tests heavy dense code-switchers on more typical stimuli may better determine the way these bilinguals store their words across the languages they speak.

To summarize, this study investigated code-switching processing in bilinguals who belong to a community where code-switching between Algerian Arabic and French is typical and dense. However, while these bilinguals differ in the amount and daily frequency of code-switching between Algerian Arabic and French they all claim that code-switching between Standard Arabic and French is not attested, not typical and find it rather odd to hear. During code-switching, bilingual speakers may anticipate a language switch. Expectancy of a language switch is more enhanced in a semantically rich context but also in a more typical context involving languages that are more frequently used in daily interactions. Anticipation of a language switch does not seem to depend solely on proficiency in the switch language. Results in the current study show that the ease of switching also depends on the habit and frequency of code-switching. These finding could be explained within the adaptive control hypothesis (Green and Abutalebi, 2013) and the control process model of code-switching (Green and Wei, 2014) which suggest that bilinguals’ daily habits of using their languages in different interactional contexts induce different habits of language control. It is worth noticing that although the switches in the stimuli are attested in speech, an ideal representative stimuli of dense code-switching would show integration of the code-switched items in Algerian Arabic base language. More representative stimuli of the heavy code-switchers would have enhanced the effect of base language and switching habits on the expectancy of a language switch. Unfortunately, such dense code-switching material cannot be tested using the current methodology because Algerian Arabic is not traditionally written. There are instances of written code-switching in social networks such as Facebook, however, it is written in French script and there is still controversy on how to spell certain Arabic sounds. A future improvement to this study would be to use auditory stimuli only with other techniques such as ERPs and eye tracking.

## AUTHOR CONTRIBUTIONS

SK designed the study, collected and analyzed the data, and wrote the manuscript; EK directed the study, and was involved in the design and data analysis, and contributed to parts of the manuscript.

## FUNDING

This study was supported by an NSF Doctoral Dissertation Research Improvement Grant awarded to SK (BCS-1451732).

## REFERENCES

- Altarriba, J., Kroll, J. F., Sholl, A., and Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: evidence from eye fixations and naming times. *Mem. Cogn.* 24, 477–492. doi: 10.3758/BF03200936
- Backus, A. (1996). *Two in One: Bilingual Speech of Turkish Immigrants in the Netherlands*. Tilburg: Tilburg University Press.
- Basnight-Brown, D. M. (2014). “Models of lexical access and bilingualism,” in *Foundations of Bilingual Memory*, eds R. R. Heredia and J. Altarriba (New York, NY: Springer), 85–107. doi: 10.1007/978-1-4614-9218-4\_5
- 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
- Bentahila, A., and Davies, E. E. (2002). Language mixing in rai music: localisation or globalisation? *Lang. Commun.* 22, 187–207. doi: 10.1016/S0271-5309(01)00026-X
- Bentahila, A., and Davies, E. E., (2001). Multiple voices: the exploitation of code switching in North African rai music. *Paper Presented at the Conference on Language(s) in North Africa: Multiple Practices, Multiple Identities and Multiple Ideologies*, American Institute for Maghrib Studies (AIMS), Tangier.
- Blackburn, A. M., and Wicha, N. (2011). The assessment of code-switching experience survey (ACSES): a new tool for assessing code-switching behavior in Spanish/English bilinguals. *Paper Presented at the Neurobiology of Language Conference*, Annapolis, MD.
- 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. (2001). Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341–345.
- 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. Psychol. Rev. Can. Psychol. Exp.* 69, 297–313. doi: 10.1037/cep0000063
- Boumans, L., and Caubet, D. (2000). “Modelling intrasentential codeswitching: a comparative study of Algerian/French in Algeria and Moroccan/Dutch in the Netherlands,” in *Arabic as a Minority Language*, ed. J. Owens (Berlin: Mouton de Gruyter), 113–180.
- Cieślicka, A. B., and Heredia, R. R. (2015). “Priming and online multiple language activation,” in *Methods in Bilingual Reading Comprehension Research*, eds R. R. Heredia, J. Altarriba, and A. B. Cieślicka (New York, NY: Springer Science+Business Media, LLC), 123–156.
- Costa, A., and Santesteban, M. (2004). Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. *J. Mem. Lang.* 50, 491–511. doi: 10.1016/j.jml.2004.02.002
- Costa, A., Santesteban, M., and Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1057–1074.
- DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504
- 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
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., and Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol. Methods* 1, 170–177. doi: 10.1037/1082-989X.1.2.170
- Duyck, W., Van Assche, E., Drieghe, D., and Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 663–679.

## ACKNOWLEDGMENT

We thank *the National School of Computer Science and the National School of Polytechnics* in Algiers/Algeria for their support.

- Elston-Güttler, K. E., Gunter, T. C., and Kotz, S. A. (2005). Zooming into L2: global language context and adjustment affect processing of interlingual homographs in sentences. *Cogn. Brain Res.* 25, 57–70. doi: 10.1016/j.cogbrainres.2005.04.007
- Gollan, T. H., and Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 640–665. doi: 10.1037/a0014981
- Green, D. W. (2011). Language control in different contexts: the behavioral ecology of bilingual speakers. *Front. Psychol.* 2:103. doi: 10.3389/fpsyg.2011.00103
- Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *J. Cogn. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377
- 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
- Grosjean, F. (2008). *Studying Bilinguals*. New York, NY: Oxford University Press.
- Gross, S. (2000). Intentionality and the markedness model in literary codeswitching. *J. Pragmat.* 32, 1283–1303. doi: 10.1016/S0378-2166(99)00102-2
- Heredia, R. R. (1997). Bilingual memory and hierarchical models: a case for language dominance. *Curr. Dir. Psychol. Sci.* 6, 34–39. doi: 10.1111/1467-8721.ep11512617
- Heredia, R. R., and Altarriba, J. (2001). Bilingual language mixing: why do bilinguals code-switch? *Curr. Dir. Psychol. Sci.* 10, 164–168. doi: 10.1111/1467-8721.001140
- Heredia, R., and Blumentritt, T. (2002). On-line processing of social stereotypes during spoken language comprehension. *Exp. Psychol.* 49, 208–221. doi: 10.1026//1618-3169.49.3.208
- Heredia, R. R., and Stewart, M. T. (2002). On-line methods in bilingual spoken language research. *Adv. Psychol.* 134, 7–28. doi: 10.1016/S0166-4115(02)80004-8
- Hernández, A. E., Bates, E. A., and Avila, L. X. (1996). Processing across the language boundary: a cross-modal priming study of Spanish-English bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 846–864.
- Kroll, J. F., and Gollan, T. H. (2014). “Speech planning in two languages: what bilinguals tell us about language production,” in *The Oxford Handbook of Language Production*, eds V. Ferreira, M. Goldrick, and M. Miozzo (Oxford: Oxford University Press).
- Kroll, J. F., van Hell, J. G., Tokowicz, N., and Green, D. W. (2010). The revised hierarchical model: a critical review and assessment. *Biling. Lang. Cogn.* 13, 373–381. doi: 10.1017/S136672891000009X
- Li, P. (1996). Spoken word recognition of code-switched words by Chinese-English bilinguals. *J. Mem. Lang.* 35, 757–774. doi: 10.1006/jmla.1996.0039
- 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, 381–390. doi: 10.1037/a0014875
- Love, T., Maas, E., and Swinney, D. (2003). The influence of language exposure on lexical and syntactic language processing. *Exp. Psychol.* 50, 204–216. doi: 10.1026//1617-3169.50.3.204
- McClelland, J. L., and O’Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 634–644.
- Meuter, R. F. I. (2009). “Language selection and performance optimisation in multilinguals,” in *Multidisciplinary Approaches to Code Switching*, eds K. de Bot, L. Iurigin, and D. Winford (Philadelphia, PA: John Benjamins Pub. Company), 27–52.
- Moreno, E. M., Federmeier, K. D., and Kutas, M. (2002). Switching languages, switching palabras (words): an electrophysiological study of code switching. *Brain Lang.* 80, 188–207. doi: 10.1006/brln.2001.2588
- Myers-Scotton, C. (2006). Natural codeswitching knocks on the laboratory door. *Biling. Lang. Cogn.* 9, 203–212. doi: 10.1017/S1366728906002549

- Poulisse, N. (2000). Slips of the tongue in first and second language production. *Stud. Linguist.* 54, 136–149. doi: 10.1111/1467-9582.00055
- Proverbio, A. M., Leoni, G., and Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: an ERP study. *Neuropsychologia* 42, 1636–1656. doi: 10.1016/j.neuropsychologia.2004.04.013
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rodriguez-Fornells, A., Kramer, U. M., Lorenzo-Seva, U., Festman, J., and Munte, T. F. (2011). Self-assessment of individual differences in language switching. *Front. Psychol.* 2:388. doi: 10.3389/fpsyg.2011.00388
- 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
- Stanovich, K. E., and West, R. F. (1981). The effect of sentence context on ongoing word recognition: tests of a two-process theory. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 658–672.
- Sunderman, G., and Kroll, J. F. (2006). First language activation during second language lexical processing: an investigation of lexical form, meaning, and grammatical class. *Stud. Second Lang. Acquis.* 28, 387–422. doi: 10.1017/S0272263106060177
- Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *J. Mem. Lang.* 27, 324–340. doi: 10.1016/0749-596X(88)90058-7
- Tabossi, P. (1996). Cross-modal semantic priming. *Lang. Cogn. Process.* 11, 569–576. doi: 10.1080/016909696386953
- 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, 1412–1431. doi: 10.1037/a0024492
- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research. *Stud. Second Lang. Acquis.* 33, 339–372. doi: 10.1017/S0272263111000015
- Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Guzzardo Tamargo, R. E., Coffman, D. L., and Gullifer, J. (2011). Costly integration of code switched utterances: when code switching becomes a language switching task. *Paper Presented at 8th International Symposium on Bilingualism*, Oslo.
- Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., and Bajo, M. T. (in press). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguist. Approaches Biling.* doi: 10.1075/lab.150101.val
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., and Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *J. Mem. Lang.* 64, 88–107. doi: 10.1016/j.jml.2010.08.006
- Van Assche, E., Duyck, W., Hartsuiker, R. J., and Diependaele, K. (2009). Does bilingualism change native language reading? Cognate effects in a sentence context. *Psychol. Sci.* 20, 923–927. doi: 10.1111/j.1467-9280.2009.02389.x
- van Hell, J. G., and de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychol. (Amst.)* 128, 431–451. doi: 10.1016/j.actpsy.2008.03.010

**Conflict of Interest Statement:** 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 © 2016 Kheder and Kaan. 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) or licensor 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.



# Gauging the Impact of Gender Grammaticization in Different Languages: Application of a Linguistic-Visual Paradigm

Sayaka Sato<sup>1,2\*</sup>, Pascal M. Gyga<sup>1</sup> and Ute Gabriel<sup>3</sup>

<sup>1</sup> Department of Psychology, University of Fribourg, Fribourg, Switzerland, <sup>2</sup> Department of Linguistics and English Language, Lancaster University, Lancaster, UK, <sup>3</sup> Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Sendy Caffarra,  
Basque Center on Cognition, Brain  
and Language, Spain  
Rachel Helen Messer,  
Oklahoma State University, USA

### \*Correspondence:

Sayaka Sato  
sayaka.sato@unifr.ch

### Specialty section:

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

**Received:** 21 May 2015

**Accepted:** 26 January 2016

**Published:** 23 February 2016

### Citation:

Sato S, Gyga PM and Gabriel U  
(2016) Gauging the Impact of Gender  
Grammaticization in Different  
Languages: Application of a  
Linguistic-Visual Paradigm.  
*Front. Psychol.* 7:140.  
doi: 10.3389/fpsyg.2016.00140

Employing a linguistic-visual paradigm, we investigated whether the grammaticization of gender information impacts readers' gender representations. French and German were taken as comparative languages, taking into account the male gender bias associated to both languages, as well as the comparative gender biases associated to their plural determiners (French: *les* [generic] vs. German: *die* [morphologically feminine]). Bilingual speakers of French and German had to judge whether a pair of facial images representing *two men* or *a man and a woman* could represent a gender stereotypical role noun prime (e.g., *nurses*). The prime was presented in the masculine plural form with or without a plural determiner. Results indicated that the overt grammaticization of the male gender in the masculine form dominated the representation of the role nouns (though interpretable as *generic*). However, the effect of the determiner was not found, indicating that only gender information associated to a human reference role noun had impacted readers' representations. The results, discussed in the framework of the *thinking-for-speaking* hypothesis, demonstrated that linguistic-visual paradigms are well-suited to gauge the impact of both stereotype information and grammaticization when processing role nouns.

**Keywords:** gender representation, gender stereotypes, grammatical gender, generic masculine, thinking-for-speaking hypothesis, bilingualism

## INTRODUCTION

The ways in which languages organize specific concepts in their linguistic systems have been found to impact how we represent information (e.g., Gennari et al., 2002; Papafragou et al., 2002). This notion, further developed as the *thinking-for-speaking* hypothesis by Slobin (1996) in his work on motion events, proposes that the encoding of concepts and events within a language acts both as a foundational and constraining structure for how verbal information is represented. Processing a specific language therefore imposes speakers to focus on particular concepts that are grammaticized within its structure, resulting in language-bound representations. As will be further discussed in this paper, bilinguals are particularly suited for testing the thinking-for-speaking hypothesis as they offer a platform to examine the extent to which comprehension mechanisms change as a function of the characteristics of the language being used (e.g., Boroditsky et al., 2003;



Fausey et al., 2010; Bylund and Jarvis, 2011). In the present study, we focus specifically on the case of gender representation during language comprehension, and argue that processing languages that grammaticize gender information in their linguistic structure will result in heightened biased representations of gender.

Recent psycholinguistic research investigating gender representation during language comprehension has shown that the presence or the lack of gender information in the linguistic structure of a language contributes to shaping distinct gender representations. For example, languages such as English, that do not systematically grammaticize gender information in their linguistic structure, encourage readers to rely on their world knowledge for gender representations (e.g., Carreiras et al., 1996; Kennison and Trofe, 2003; Oakhill et al., 2005; Reynolds et al., 2006; Pyykkönen et al., 2010). Reading about person references such as *nurse* will generate inferences about the possible gender of the depicted person, with gender stereotypes acting as a primary source for representation (e.g., Banaji and Hardin, 1996; Carreiras et al., 1996; Kennison and Trofe, 2003; Cacciari and Padovani, 2007; Kreiner et al., 2008). Banaji and Hardin (1996), for example, showed that participants' judgments to the target stimuli (Experiment 1: judge whether the target was a male or female: *he vs. hers*; Experiment 2: judge whether the target was a pronoun or not: *she vs. do*) following either a gender stereotypical (e.g., nurse, mechanic) or gender definitional (e.g., mother, king) prime was found to be responded to faster when there was a gender congruency between the prime and target stimuli. Oakhill et al. (2005) further substantiated these effects of gender priming with a series of lexical priming experiments. Participants in their study were faster to accept word pairs consisting of a stereotypical role noun (e.g., surgeon) and kinship term (e.g., brother, sister) as referring to the same person in cases when the words were gender congruent. Activating such stereotyped gender inferences has been found to be immediate and robust among English readers, demonstrating that such role nouns may prime a specific stereotypical gender even if morphological or grammatical information may not compel readers to do so (e.g., Carreiras et al., 1996; Kennison and Trofe, 2003).

These representation tendencies, however, are not readily generalizable for readers of grammatical gender languages such as French or German, where stereotypical gender is only one of the two possible sources contributing to the construction of gender representations. In these languages, gender is also integrated as part of their grammatical structure. Grammatical gender thus classifies a specific gender category to all nouns (e.g., masculine, feminine, and neuter in the case of German). This gender feature, when marked on person references, commonly corresponds to the biological gender of the referent (i.e., masculine = man, feminine = woman)<sup>1</sup>, constraining its

language users to consistently monitor gender information at both grammatical and semantic levels. A fundamental claim made by researchers is that the interaction between these two sources of information (i.e., stereotypical and grammatical) during the processing of role nouns is complex, and that the mechanisms for representing gender information are not always straightforward (e.g., Irmen, 2007; Garnham et al., 2012; Gygas et al., 2012; Esaulova et al., 2013).

The complexity of this interaction is rendered by the fact that gender information associated to its surface form does not necessarily coincide with its intended semantic connotations. For instance, when considering the masculine form, there is a discrepancy between form and meaning. Whereas role nouns such as *infirmières*<sub>Feminine</sub> [nurses] marked in the feminine grammatical form refer unambiguously to female nurses, the masculine form (*infirmiers*<sub>Masculine</sub>) can refer exclusively to men (i.e., only male nurses) or it may refer to a group composed of both male and female persons (i.e., *generic* interpretation). Readers are presented with a challenge to disambiguate the intended interpretation of the masculine form. It has been argued that its surface forms naturally emphasize the association to the male gender, inevitably prompting a male-specific interpretation (e.g., Gygas et al., 2012). Gygas et al. (2012), for example, adapting Oakhill et al.'s (2005) paradigm in French, found that when participants were instructed to decide whether the person represented by a kinship term in pairs such as *tante* [aunt] – *infirmiers*<sub>Masculine</sub> [nurses] could belong to a group represented by the second noun (always in the grammatical masculine plural form), they responded positively more often and faster when the kinship term was a man, indicating a male dominant representation. The authors concluded that the generic interpretation could only be activated through active processes, yet the male-specific interpretation was always passively activated (i.e., without control). Most studies using on-line (e.g., Gabriel and Gygas, 2008; Gygas et al., 2008) and off-line (Stahlberg et al., 2001; Braun et al., 2005) tasks concur on the male-specific impact of the masculine form. Crucially, however, this male bias effect persisted even when gender stereotypicality violated the grammatical gender information (as seen in *infirmiers*: female stereotype, masculine grammar), leaving the effect of stereotype information unclear.

In German, additional grammatical cues associated to its plural determiner (*die* [the]) and pronoun (*sie* [they]) have been investigated, especially in conjunction with possible female biases. In a study investigating gender representation in German, Rothermund (1998) found an unexpected reduction of the male bias when participants conducted a recognition task after reading texts including plural masculine references (*die*<sub>plural</sub> *Studenten* [the students]). The male-attenuating effect was attributed as being triggered by the plural determiner *die* which shares the same surface form as the singular feminine determiner *die* [*the* – singular – feminine]. Garnham et al. (2012) also showed a male attenuated effect (or an additive female effect) when presenting the German plural pronoun *sie* (i.e., *they* – also feminine-equivalent) in a sentence judgment task examining the interpretations of masculine role nouns. When the same was done in French, however, the masculine

<sup>1</sup>Note that grammatical gender does not determine gender for all human references and the relation is mutual. For instance, there are special cases such as bi-gender role nouns where the sex of the referent determines the grammatical gender (e.g., *artiste* [artist – French] can be either masculine or feminine depending on the gender of the person). Additionally, there are also examples of epicenes where a single gender can refer to both the biological sex (e.g., *secrétaire*<sub>Feminine</sub> is always feminine irrespective of whether it refers to a male or a female).

pronoun *ils* [they – masculine *specific* or *generic*] did not have a male amplifying effect despite its male association. The authors argued that although cumulating grammatical cues do not augment male biases, combinations of male and female-equivalent grammatical cues may distract readers from activating male specific representations. To our knowledge, when looking strictly at determiners, only one study (e.g., Gygas et al., 2008) has generated specific hypotheses as to the impact of the definite plural determiner *die* in German, yet its female-bias effect (as shown by Rothermund, 1998) was never clearly replicated.

The studies discussed here demonstrate how grammaticized information influences readers' comprehension processes. Grammatical gender languages work in a top-down manner, constraining their users to consistently monitor gender both on grammatical and semantic levels. If, as suggested by the thinking-for-speaking hypothesis, information grammaticized in languages prompts readers' gender biases, which in turn anchor their representations, these regularities should also become evident on their representations. If this were the case, it is reasonable to assume that readers of more than one language may switch representations as they change languages. This notion is further developed in this study by looking particularly at bilinguals where the language biases of each of the bilingual's languages should surface on their representations. Sato et al. (2013) followed this line of logic and investigated in a sentence-based paradigm, whether English-French bilinguals would construct different representations according to their first (L1) and second (L2) language. They presented English and French bilingual participants with sentence primes including role nouns with stereotypical gender (e.g., female: *nurses*, male: *politicians*, neutral: *pedestrians*). Participants judged the plausibility of target sentences including a gender reference (e.g., *some men*, *some women*) to be a sensible continuation of the prime. The results indicated that switching language was also accompanied by changes of biases in mental representations of gender, with English eliciting stereotyped representations and French male-biased representations triggered by the masculine form. Importantly, participants' L2 proficiency was found to be a good indicator of the extent of the representation switch between L1 and L2.

In the present study, we followed Sato et al.'s (2013) study and investigated the effects of stereotypes and linguistic encodings of gender on the representation of person reference role nouns. French and German were taken as comparative languages, provided that they were both marked with grammatical gender. This made them ideal candidates to test thinking-for-speaking effects, as opposed to English, which lacks systematic grammaticization of gender. Characteristics surfacing on representations when processing French and German should essentially reflect the impact of how linguistic encoding contributes in shaping gender representations. Additionally, despite their common usage of the masculine form to denote a generic interpretation, gender associations linked to the plural determiners differ in the two languages. As argued by Rothermund (1998) and Gygas et al. (2008), the German determiner *die* [the – plural] shares the same surface structure as

the singular feminine determiner *die* [the – singular – feminine], and should contribute to a female additive bias when presented with a role noun in the masculine form. In contrast, the French plural determiner *les* [the – plural] corresponds to both feminine and masculine nouns as they have a single morphological realization (i.e., gender syncretism: Corbett, 1991) and therefore should not enhance any additional gender information. If in the present study we are able to observe differences in gender biases between French and German representations, it should provide more compelling evidence as to the impact the grammaticization of language has on our conceptualization of gender information.

To test these effects, we employed a combined linguistic-visual paradigm. This paradigm was intended to provide a more sensible experimental framework to address the immediacy of gender activation. While a handful of studies have examined gender representation processes employing a lexical-based paradigm (Banaji and Hardin, 1996; Oakhill et al., 2005; Cacciari and Padovani, 2007; Gygas et al., 2012; Siyanova-Chanturia et al., 2012), none have directly addressed the impact of the use of the masculine form, or of role noun determiners. Studies investigating these effects have approached the issue with a sentence comprehension task, applying anaphor resolution paradigms that were dependent on the detection of semantic and syntactic inconsistencies in comprehension. These tasks therefore did not strictly speak to the immediacy of the activation of such surface-level grammatical cues, and discursive contextual elements may have interfered with stereotype activation or with the accessing of signals during activation. More importantly, although moderate, some effects of stereotype have been observed, indicating that teasing apart these effects in a linguistic context has been complex.

For instance, Esaulova et al. (2013) found a subtle effect of stereotypical gender in German. In their experiment (Experiment 1), participants were presented with sentences composed of an anaphor (e.g., *er* [he]) and a stereotyped role noun (e.g., *der Elektriker* [the electrician]) as an antecedent while their eye movements were recorded. Although comprehension difficulty was most prominent when the anaphor did not agree grammatically with its antecedent, as illustrated by most eye-tracking measures, sentence processing was also influenced by the role nouns' stereotypicality, as demonstrated in the late measures only (e.g., regression path on the pronoun region and total fixation path on the role noun). Following the aforementioned Banaji and Hardin's (1996) experiments, Cacciari and Padovani (2007) and Siyanova-Chanturia et al. (2012) also reported stereotype effects in Italian, a grammatical gender language. They found that when a pronoun (e.g., *lui* [he] or *lei* [she]) was primed by a bi-gender role noun (a noun that can vary in grammatical gender as a function of biological gender, as in *insegnante* [a female/male teacher]), participants were particularly slow to decide whether the pronoun was masculine or feminine when primed by a counter-stereotypical role noun. Additionally, Carreiras et al. (1996), in a self paced reading task (Experiment 2), showed that Spanish participants reading was delayed when a role noun (e.g., the carpenter) was written in a grammatical form that mismatched its stereotypicality (e.g., *La*

*carpintera*<sub>Feminine</sub> [the female carpenter] or *El enfermero*<sub>Masculine</sub> [the male nurse]).

In sum, most studies have shown a strong impact of grammatical gender, with some authors claiming that grammatical gender had only overshadowed stereotype effects (e.g., Irmen, 2007; Esaulova et al., 2013; Reali et al., 2015). Although, the impact of grammatical cues seems central in representation processes, the reasons for the overriding effects of grammatical cues over gender representations have not been clearly shown. We therefore explore the possibility that the prevalence of male representations in grammatical gender languages (and the lack of stereotype effects) may have well been prompted by the very nature of the paradigms being employed, provided that both the prime and target stimuli were verbal stimuli. The use of verbal target stimuli, maintaining a close link with its verbal prime, may have resulted in mental representations that reflected only and merely linguistic activations. It could be that processing both prime and target stimuli in a verbal context may constrain readers to over-monitor grammatical and syntactical properties. This monitoring in turn may enhance the signal of a representation based on linguistic cues (i.e., toward a male bias in gender-marked languages). In contrast, linguistic-visual paradigms have been found to be effective in gauging effects of gender priming. Studies in social psychology have shown that gender priming may be observed by presenting gender associated words (e.g., Kawakami and Dovidio, 2001: stereotypical traits; Lemm et al., 2005: words with gender-specific suffixes and role nouns) followed by picture targets that required participants' judgments. For instance, Lemm et al. (2005) showed that although past studies indicated a weaker priming effect when using cross-modal paradigms, the gender priming effects found in their study were still large. Consequently, this approach may indeed be well-suited to gauge the subtle stereotype effects we seek to explore.

In our task, stereotypical role nouns in the masculine plural form, either with or without a plural determiner, served as gender primes in German or in French. Participants had to make judgments as to whether a visually presented pair of faces (male pairs or mixed pairs of faces composed of a woman and a man) that followed could represent the preceding prime. The composition of face pairs represented the possible interpretations that the role noun in the masculine form holds (i.e., a male specific or a generic interpretation). We expected to replicate the male bias demonstrated in previous findings (i.e., facilitated responses to male pairs of faces), and intended to explore the influence of stereotype information. Specifically, an attenuated male bias was expected in the female and possibly the neutral stereotyped conditions. Importantly, we also expected that the determiner *die* in German would attenuate this potential male bias arising from the masculine form of the role noun, whereas French rolenouns would retain the male bias. Finally, the experimental task was carried out in participants' L1 and L2 to examine any representational shift that would be prompted by the regularities of each language. For participants' L2, we also took L2 proficiency into account, as measured by a L2 C-test. We expected shifts of representations to be influenced by L2 proficiency (as in Sato et al., 2013).

## MATERIALS AND METHODS

### Participants

#### Native German Group

Fifty Caucasian German-speaking students from the University of Fribourg (Switzerland) participated in the experiment for course credits. All participants were native speakers of German whose L2 was French (mean age: 22, mean start age of French acquisition: 9.4 years, mean number of schooling of French as L2: 7.2 years). Forty-one participants were women<sup>2</sup>.

#### Native French Group

Fifty-one Caucasian French-speaking students from the University of Fribourg participated in the experiment for course credits. All participants were native speakers of French whose L2 was German (mean age: 22, mean start age of German acquisition: 7.5 years, mean number of schooling of German as L2: 9.2 years). Thirty-nine participants were women.

### Materials

#### Prime Role Nouns

Thirty-six gender stereotypical role nouns were selected as primes for the experiment (see **Table 1**). These role nouns were taken from Gygax et al. (2008), all of which were normed and tested for gender stereotypicality in Gabriel et al. (2008) in both German and French. Role nouns were female (e.g., nurses [*Krankenpfleger/infirmiers*]), male (e.g., bosses [*Arbeitgeber/patrons*]) or neutral (e.g., pedestrians [*Spaziergänger/promeneurs*]) in stereotype. To ensure that both female and male stereotyped role nouns were similarly judged as prototypical exemplars of their respective stereotype, we inverted ratings to female stereotypes (i.e., new rating = 100 - initial ratings), and conducted a *t*-test to ensure that both were similarly judged. As expected, both were similar in both languages,  $t_{French}(22) = 0.23, p = 0.82$ , and  $t_{German}(22) = 0.47, p = 0.64$ .

#### Target Face Pairs

The face pairs were created with the face modeling software FaceGen Modeler program version 3.1.4 (Singular Inversions Inc, 2004; Toronto). A total of 30 male and 30 female Caucasian faces with neutral expressions were created. They all had neutral expressions and the crown area of the faces were removed in order to eliminate possible biases associated with certain hairstyles evoking gender-biased information.

Twenty-one participants (14 women and seven men who did not participate in the main experiment) participated in the first norming phase by rating the gender typicality of all 60 faces on a 7-point scale (very masculine = 1, very feminine = 7) on a paper-pencil administrated questionnaire. Presentation order of the faces was randomized for each participant. Only faces that were clearly rated as female (i.e., average score > 5) or male (i.e., average score < 3) were selected for the experiment. Twenty-four female faces ( $M = 5.72$ ,

<sup>2</sup>As past studies on gender representation (e.g., Gygax and Gabriel, 2008; Gygax et al., 2008; Garnham et al., 2012) did not find effects of participants' gender in reading tasks, we did not balance the gender sample of our participants.

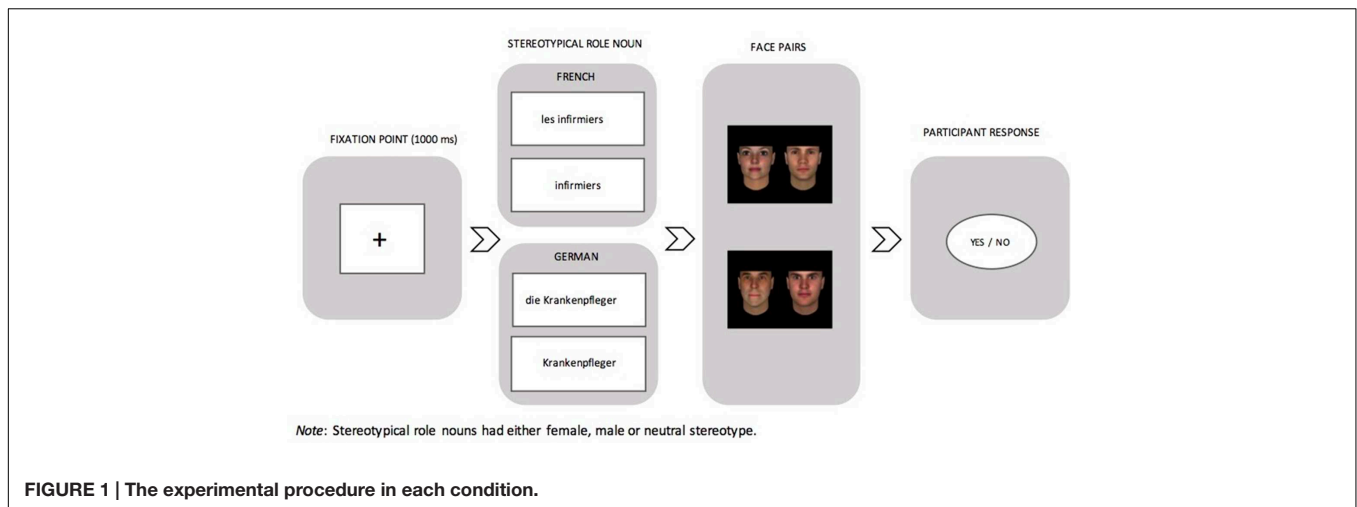
**TABLE 1 | Role nouns from Gabriel et al. (2008) and their corresponding gender proportion and standard deviations (in parentheses) for each stereotype.**

English	German	% (SD)	French	% (SD)
Spies	Spione	67 (15)	Espions	74 (17)
Golfers	Golfspieler	68 (14)	Golfeurs	73 (16)
Politicians	Politiker	69 (11)	Politiciens	72 (13)
Police officers	Polizisten	69 (10)	Policiers	70 (13)
Statisticians	Statistiker	72 (12)	Statisticiens	74 (15)
Bosses	Arbeitgeber	72 (12)	Patrons	74 (16)
Computer specialists	Informatiker	79 (11)	Informaticiens	67 (22)
Surgeons	Chirurgen	75 (12)	Chirurgiens	75 (14)
Technicians	Techniker	78 (14)	Techniciens	75 (14)
Engineers	Ingenieure	78 (11)	Ingénieurs	74 (14)
Physics students	Physikstudenten	81 (11)	Etudiants en physique	67 (28)
Pilots	Flieger	76 (13)	Aviateurs	74 (17)
<b>Mean</b>		74 (5)		72 (3)
Singers	Sänger	45 (8)	Chanteurs	48 (9)
Pedestrians	Spaziergänger	46 (8)	Promeneurs	52 (13)
Cinema goers	Kinobesucher	49 (6)	Spectateurs de cinéma	50 (5)
Concertgoers	Konzert-Zuhörer	47 (7)	Auditeurs de concert	51 (10)
Schoolchildren	Schüler	48 (5)	Ecoliers	53 (13)
Spectators	Zuschauer	41 (8)	Spectateurs	51 (5)
Neighbours	Nachbarn	50 (5)	Voisins	50 (8)
Swimmers	Schwimmer	50 (9)	Nageurs	50 (10)
Tennis players	Tennispieler	52 (7)	Joueurs de tennis	54 (8)
Authors	Autoren	52 (9)	Auteurs	54 (8)
Musicians	Musiker	50 (9)	Musiciens	59 (13)
Skiers	Skifahrer	53 (8)	Skieurs	55 (9)
<b>Mean</b>		49 (3)		52 (3)
Beauticians	Kosmetiker	11 (8)	Esthéticiens	18 (20)
Birth attendants	Geburtshelfer	11 (19)	Assistants maternels	18 (18)
Fortune tellers	Wahrsager	24 (16)	Diseurs de bonne aventure	28 (27)
Cashiers	Kassierer	27 (16)	Caissiers	24 (15)
Nurses	Krankenpfleger	24 (11)	Infirmiers	30 (11)
Hairdressers	Coiffeure	21 (11)	Coiffeurs	38 (25)
Psychology students	Psychologiestudenten	25 (29)	Etudiants en psychologie	33 (10)
Dieticians	Diätberater	27 (15)	Dietéticiens	37 (22)
Dressmakers	Schneider/Näher	23 (13)	Couturiers	40 (32)
Dancers	Tänzer	33 (12)	Danseurs	29 (14)
Sales assistants	Verkäufer	33 (14)	Vendeurs	37 (13)
Social workers	Sozialarbeiter	41 (14)	Assistants sociaux	33 (15)
<b>Mean</b>		25 (9)		30 (7)

All role nouns are presented in the plural form as was in the experiment.

$SD = 0.33$ , range = 5.43–6.3) and all thirty male faces ( $M = 1.58$ ,  $SD = 0.26$ , range: 1.23–2.47) were retained. The average ratings of the female faces [ $t(23) = 25.27$ ,  $p < 0.001$ ;  $M_{\text{difference}} = 1.72$ ] and male faces [ $t(29) = -50.17$ ,  $p < 0.001$ ;  $M_{\text{difference}} = -2.42$ ] were significantly different from the scale midpoint (i.e., 4), with the difference being bigger for male faces than for female faces. We deemed this imbalance in deviation from midpoint non-problematic for the purpose of our study, as our main focus was on assuring to select non-ambiguous faces.

The 54 faces were then combined to make *male* and *mixed pairs of faces* (see **Figure 1** for an example of a presented pair of faces). Female pairs of faces were not constructed for the experiment, as the interpretation of the presented masculine forms could not be grammatically interpreted as being female-specific (i.e., represented by female pairs of faces). More importantly, these female pairs of faces were avoided based on findings by Gygas and Gabriel (2008) who demonstrated that the presentation of both feminine and masculine forms in the same experiment directs readers



**FIGURE 1 |** The experimental procedure in each condition.

toward a stronger male-specific representation of the masculine form. Female faces for mixed pairs were always presented on the left in order to avoid a male preferred response according to a possible left-side bias, illustrated in past studies using response scales in left-to-right languages (e.g., Gabriel et al., 2008). All pairs of faces were comprised of different faces.

A second norming phase was conducted in order to ensure that male and mixed pairs of faces were not processed differently due to perceptual properties that we had not foreseen. In this pilot experiment, our experimental pictures were presented on a computer screen running Experiment Builder (SR Research) to another group of 27 participants (25 women and six men who had not participated in the first norming phase). Their task was to decide, on two blocks of trials, whether the presented pairs of faces were *of the same sex* in one block or *of different sex* in the other block, by indicating their responses with a *yes* or *no* button press. The block order was inverted for half of the participants. A repeated-measure ANOVA on correct response times (i.e., 94% of the data) showed no main effect of *block*,  $F(1,26) < 1$ , *ns.*, no main effect of *faces*,  $F(1,26) = 3.18^3$ , *ns.*, and no interaction,  $F(1,26) = 1.75$ , *ns.*, confirming the homogeneity of our experimental target stimuli in terms of perceptual properties.

## L2 Proficiency Assessment

Participants' L2 proficiency levels were operationalized by their performance scores on a given C-test (as done in Sato et al., 2013). Commonly in a C-test, participants are given several distinct passages in which the second half of every other word is deleted except for the first and last sentences. The task is to restore the blanks in the allocated time. This procedure was developed as an effective measurement substituting cloze tests that were used in earlier years, and in recent years, has been frequently encouraged

as a measure for language proficiency (Grotjahn et al., 2002; Eckes and Grotjahn, 2006).

In fact, C-tests have been shown to be highly correlated with standardized tests (e.g., Studienkollegs in German: Grotjahn and Allner, 1996; TOEFL in English: Hastings, 2002; the five competencies of the Test de Connaissance du Français: Reichert et al., 2010). We employed the German C-test offered by onDaF<sup>4</sup> to test German proficiency. Score ratings on this test are considered equivalent to the Common European Framework of Reference for the levels A2 to C1. French proficiency was evaluated with Coleman's (1994) C-test. Four texts were chosen from each original version and 20 min were allocated to complete the task.

## Role Noun Translation Task

To verify whether participants correctly identified the role nouns presented in L2, a role noun translation task was conducted after the experimental trials. Participants were asked to provide a translation for each presented role noun in their L1.

## Design and Procedure

The experimental task was conducted first in L1, followed by the task in L2 to minimize any data contamination during the processing of a less dominant language<sup>5</sup>. Two experimental lists were created to ensure that a role noun would not appear in both languages for a given participant. The two lists were symmetrically different, in that if a role noun appeared in French in List 1, in List 2, it would appear in German. To avoid an imbalance of gender stereotypicality between languages, role nouns of similar strength of stereotype were always allocated to each language (see **Table 1**). Each list consisted of six female, six male and six neutral role nouns per language, resulting in 36 critical role nouns per list, with each role noun appearing only in either French or German. Each role noun was presented

<sup>4</sup>www.ondaf.de

<sup>3</sup>If anything, participants were slightly faster (by 36 ms) to respond to mixed pairs of faces than to male pairs ( $p = 0.08$ ).

<sup>5</sup>Experiment order for participants' L1 and L2 was not randomized as Sato et al. (2013) did not find any experimental order effects based on language.

four times per participant (Oakhill et al., 2005; cf. Gygax et al., 2012, for a similar procedure): twice with a determiner (once followed by male pairs, once by mixed pairs of faces), and twice without. All experimental items were intended to elicit a *yes* response.

To trigger *no* responses, twenty filler role nouns that had a gender association by definition (e.g., grandmother: *Großmütter/grand-mères*) were included. Half of the filler role nouns were male by definition, whereas the other half were female. These filler primes were also presented four times with their respective determiner allocations and face pairs. As these nouns were not ambiguous in terms of gender, including them prevented participants from responding *yes* throughout the experimental task without truly processing the role nouns and the target stimuli.

The study was accepted by the Ethics Committee at the Department of Psychology of the University of Fribourg and conformed to relevant regulatory standards. All participants were granted informed consent. For each experimental trial, participants were first presented with a gender stereotypical role noun prime following a fixation point (1000 ms). The role noun was presented in the masculine plural form either in conjunction with a plural definite determiner (e.g., *die Ingenieure/les ingénieurs* [the engineers]) or without (e.g., *Ingenieure/ingénieurs* [engineers]). Participants were instructed to press the *yes* button after having read the presented role noun, which prompted the presentation of a picture of a pair of faces. Their task was to judge as quickly as possible with a *yes/no* button press whether the presented target face pairs could represent the prime role noun that appeared prior to the faces (see **Figure 1** for the procedure). Filler trials, which were randomized among experimental trials followed the same procedure, and the role nouns within them were also presented either with or without a determiner.

The experiment was run on a Power Macintosh 4400 with the Psycscope software (Cohen et al., 1993) connected to a button box to provide millisecond accuracy responses. Two buttons were labeled, one “*Ja*” (yes) and the other “*Nein*” (no) for German-speaking participants and “*Oui*” (yes) and “*Non*” (no) for French-speaking participants. Items were presented on a computer screen and the “*Ja/Oui*” button was always pressed by the participant’s dominant hand. All participants were individually tested in a quiet room, with instructions being given in their respective native languages. They underwent a practice session in their L1 with four items to familiarize themselves with the task and procedure.

After the main experimental task, three paper-based post-tests were conducted. First, participants were given a C-test in their respective L2. Following the C-test, participants were requested to assess their L2 competence in terms of their listening, reading, writing and speaking abilities in the L2 and to indicate the years and age of L2 acquisition by means of a self-administered questionnaire. Finally, the role noun translation task was given to the participants to ensure they had properly processed the critical items.

## RESULTS

We conducted analyses on both participants’ binary responses (*yes/no*) to the facial images and their response times for *yes*-responses (i.e., accepting the faces). Based on the results of the role noun translation task conducted after the main experimental task, items in the L2 that were frequently unknown to each language group (fewer than 10% of the participants were able to provide a correct translation) were omitted from the analyses (*Schneider* [dress makers] and *Wahrsager* [fortune tellers] were removed from L2-French participants’ data and *diseurs de bonne aventure* [fortune tellers] from L2-German participants’ data). Mixed-effects logistic regression was used to model the binary outcome variable (*yes/no* responses), and linear mixed-effects regression was used to model participants’ positive response times. Mixed-effect models provide a means to perform analyses that account for missing values and to avoid the language-as-a-fixed-effect fallacy (Brysbaert, 2007). All analyses were conducted using the R software (R Core Team, 2013), with the *glmer* and *lmer* function from the *lme4* package (Bates et al., 2014). As suggested by Barr et al. (2013) a model with a maximal random factor structure was adopted. Random intercepts and slopes were varied for participants and items in order to account for the variance in performance created by the factors (Baayen et al., 2008). Random slopes were eliminated if their removal did not result in a significant amelioration of the model or if the model did not converge. All predictors for fixed effects were sum coded (+1, -1) and were entered by step-wise forward selection to an initial null model. Given that participants’ L2 *proficiency* was expected to predict general performance in the L2, the *proficiency* predictor, as measured by C-test scores, was centered and entered as a covariate in the null model, which included only random effects. Analyses for each language group were conducted separately as in Sato et al. (2013), given that we expected different variances in the C-test scores. Indeed, C-test difficulty has been found to vary according to various factors such as the language of the C-test, text type or deletion pattern (Sigott and Köberl, 1996).

Log-likelihood ratio tests were used to determine the adequacy of including each predictor in the model. A more complex model including the predictor in question was compared to a simpler model without the inclusion of the predictor. If its integration significantly improved the model, the predictor was retained within the model. The predictors tested in the models were *face pairs* (male vs. mixed pairs of faces), *stereotype* (female vs. male vs. neutral), *task language* (German vs. French) and *determiner* (without determiner vs. with determiner).

### Responses to Facial Targets

Participants’ binary choices were modeled in a mixed logit model to predict the likelihood that participants would accept a face pair presented to them after a particular role noun prime. For both language groups, the first model that followed the null model tested the effects of the masculine form and of stereotype by introducing *face pairs*, *stereotype* and their interaction to the null model. For both groups, the inclusion of these predictors significantly improved the model fit (Native German group:

$\chi^2 = 205.36$ ,  $df = 5$ ,  $p < 0.001$ ; Native French group:  $\chi^2 = 150.8$ ,  $df = 9$ ,  $p < 0.001$ ). The second model proceeded to test whether the effect of the German determiner impacted the interpretation of the presented prime by adding the main effects of *task language* and *determiner*, and importantly, all interactions between *face pair*, *task language* and *determiner*. While this led to a significant improvement of the model for the native German group ( $\chi^2 = 75.8$ ,  $df = 6$ ,  $p < 0.001$ ), the model failed to converge for the native French group. Therefore, only main effects for *task language* and *determiner*, as well as the interaction between *task language* and *face pairs* were introduced into the model, indicating an improvement ( $\chi^2 = 31.69$ ,  $df = 5$ ,  $p < 0.001$ ). As for the random structure, the final model for the native German group included random slopes for *determiner* at the item level. The model for the native French group included random slopes for *face pair* at the item level and *stereotype* and *task language* at the participant level. Both models indicated a variance inflation factor less than 1.5, indicating that collinearity was not an issue.

### Native German Group

The results showed significant main effects of *face pairs* and *stereotype* which were qualified by a significant interaction. Overall, the likelihood of a positive response was substantially higher for male pairs of faces than for mixed pairs of faces ( $b = 0.47$ ,  $SE = 0.04$ ,  $p < 0.001$ , odds ratio = 2.56). Consistent with our predictions, the *face pairs*  $\times$  *stereotype* interaction revealed that this preference for male pairs of faces was especially pronounced when they followed role nouns with male stereotype, compared to when they followed role nouns with neutral stereotype ( $b = 0.43$ ,  $SE = 0.06$ ,  $p < 0.001$ , odds ratio = 2.36) or female stereotyped role nouns ( $b = 0.68$ ,  $SE = 0.2$ ,  $p < 0.001$ , odds ratio = 3.89; see **Figure 2**).

The model also revealed main effects of *determiner* and *task language*, indicating that the likelihood of receiving positive responses was higher if face pairs were preceded by role nouns with an article than when presented without the article ( $b = -0.19$ ,  $SE = 0.05$ ,  $p < 0.001$ , odds ratio = 0.68). Face pairs were also more likely to be responded to positively when they were presented with role nouns in participants' L2 French than when presented with role nouns in their dominant L1 German ( $b = -0.09$ ,  $SE = 0.04$ ,  $p < 0.05$ , odds ratio = 0.83). Contrary to our predictions, these two predictors did not interact, which would have supported the effect of the German determiner *die*. However, a *face pairs* by *task language* interaction surfaced, indicating that male pairs of faces were more likely to be responded to positively when preceded by a role noun in participants' L1 German than when preceded by a role noun in their L2 French ( $b = 0.24$ ,  $SE = 0.04$ ,  $p < 0.001$ , odds ratio = 1.62).

### Native French Group

As was the case for the German group's responses, the analysis of the French sample revealed a significant main effect of *face pairs* and a marginal significant effect of *stereotype* further qualified by their significant interaction. The likelihood of accepting face pairs was again higher for male pairs of faces than for mixed pairs of faces ( $b = 0.31$ ,  $SE = 0.05$ ,  $p < 0.001$ , odds ratio = 1.86). The

interaction revealed that the likelihood for participants to accept men pairs of faces was again substantial when they followed male stereotyped role nouns than when they followed neutral ( $b = 0.42$ ,  $SE = 0.07$ ,  $p < 0.001$ ,  $p < 0.01$ , odds ratio = 2.31) or female stereotyped role nouns ( $b = 0.75$ ,  $SE = 0.12$ ,  $p < 0.001$ , odds ratio = 4.48; see **Figure 2**).

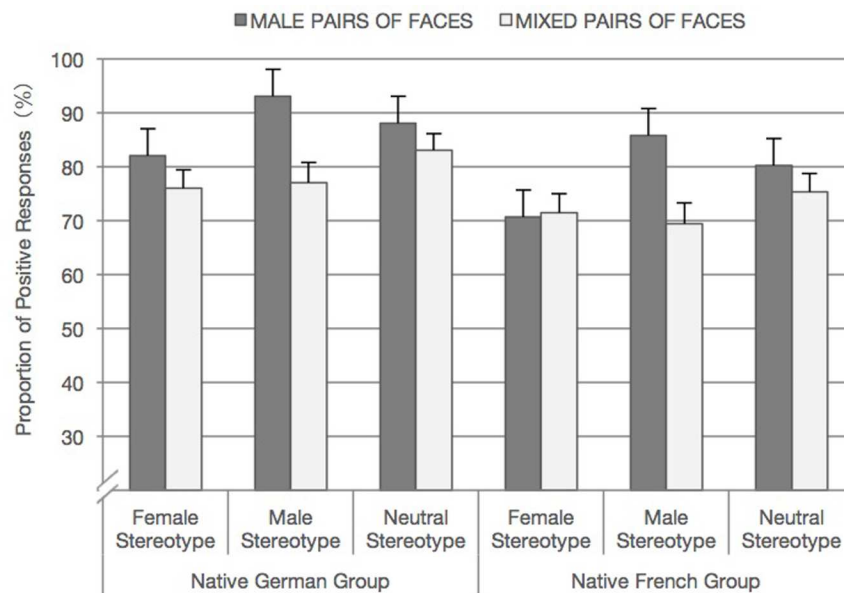
While the model revealed a main effect of *determiner* indicating that role nouns without a determiner triggered greater positive responses ( $b = 0.08$ ,  $SE = 0.03$ ,  $p < 0.05$ , odds ratio: 1.17), no interactions involving this predictor were observed. Finally, as was the case for the native German group, a significant *face pairs* by *task language* interaction indicated that responses to accept men pairs of faces was greater in the dominant L1 French than in the L2 French ( $b = -0.08$ ,  $SE = 0.03$ ,  $p < 0.05$ , odds ratio = 0.85).

## Positive Response Times to Facial Targets

Overall, both groups responded above chance level to accept facial targets (native German group: *yes* responses = 83%, *no* responses = 17%; native French group: *yes* responses = 75%, *no* responses = 25%), although these items were intended to elicit positive responses. Only reaction times to these positive responses were subject to analyses. Response times that were 2.5 standard deviations above or below the participant's mean were replaced by their cut-off values (3.5%). Following the analyses of participants' responses to face pair targets, the effects of the masculine form and stereotype were examined by introducing the main effects of *face pairs*, *stereotype* and their interaction to the null model. There were significant improvements to the models for each language group (Native German group:  $\chi^2 = 157.67$ ,  $df = 5$ ,  $p < 0.001$ ; Native French group:  $\chi^2 = 210.94$ ,  $df = 5$ ,  $p < 0.001$ ). The second model then added the main effects of *task language* and *determiner* and all interactions between *face pair*, *task language* and *determiner* in order to test the impact of the German determiner. The additions of these predictors resulted in an improvement of the models for the native German group ( $\chi^2 = 123.25$ ,  $df = 6$ ,  $p < 0.001$ ) but not for the native French group. Given that none of the effects introduced in the second model were significant, the initial model was retained as the final model. For the native German group, the random structure for the final model included random slopes for *face pairs*, *determiner* and *task language* at the item level and *face pairs*, *stereotype*, *task language* and *determiner* on the participant level. The model for the native French group included random slopes for item level and *face pairs*, *stereotypes* and their interaction for participant level. Collinearity was not an issue given that both models indicated a variance inflation factor less than 1.9.

### Native German Group

Consistent with the analyses for participants' responses, the final model showed significant main effects of *stereotype* and *face pairs*, which were qualified with their significant interaction. Male pairs of faces (825 ms) were responded to significantly faster than mixed pairs of faces (995 ms);



**FIGURE 2 | Mean proportion of positive responses to accept facial images for each native group in each stereotype condition (independent of task language).** Error bars indicate standard errors.

$b = -108.5$ ,  $SE = 29.04$ ,  $t = -3.73$ ) confirming the male bias in past studies. This male bias was more prevalent when role nouns preceding facial targets were of male stereotype than when they were neutral ( $b = -47.64$ ,  $SE = 13.89$ ,  $t = -3.43$ ) or female ( $b = -60.47$ ,  $SE = 24.27$ ,  $t = -2.49$ ) stereotype. No main effects or interaction effects including *determiner* were found, but a significant *task language* effect indicated that participants were faster to respond in their L2 French (824 ms) than in their L1 German (991 ms).

### Native French Group

The model revealed significant main effects of *stereotype* and *face pairs* that were further qualified by their significant interaction. Participants responded to male pairs (870 ms) of faces significantly faster than to mixed pairs (1009 ms) of faces ( $b = -75.83$ ,  $SE = 15.88$ ,  $t = -4.77$ ) confirming a male bias. This effect was stronger for responses to male pairs of faces following male stereotyped role nouns than when following neutral ( $b = -52.71$ ,  $SE = 13.62$ ,  $t = -3.87$ ) or female ( $b = -77.92$ ,  $SE = 23.78$ ,  $t = -3.28$ ) stereotyped ones. Contrary to our initial expectations, no effects including *determiner* were significant.

## DISCUSSION

The aims of the present study were twofold. First, we aimed to evaluate how linguistic encoding of gender in different languages shape and shift gender representations. Bilinguals of German and French were tested to assess the rather inconclusive effects of a female bias associated to the German determiner *die* (gender non-specific in the plural, but sharing the same surface form as

the feminine singular determiner). Although the activation of a male bias was anticipated, the presence of an additional female association (i.e., *die*) was expected to attenuate this male bias in German. The second goal was to provide more compelling evidence of main and interaction effects when both stereotypical and grammatical gender information are available during the processing of role nouns. It has been argued that the impact of gender stereotype information has often been overshadowed by grammatical gender information in past studies, resulting in some uncertainty as to how stereotype information actually influences the interpretation of the masculine form. While past studies have relied on verbal targets, we argue that these tasks may have reinforced the grammatical and morphological cues being tested. Such an impact may have resulted in strong, yet less generalizable grammatical-based representations. In order to overcome these issues, a new experimental approach using visual targets was suggested.

Overall, we found a consistent main effect of *face pairs* for both of our groups, where responses to male pairs of faces were facilitated over mixed pairs of faces. This facilitation reflects the general ease in interpreting role nouns in the masculine form as being male-specific rather than generic. Although the surface form of the masculine grammar can theoretically be detached from its semantic association *masculine = men*, it nonetheless boosted the activation of semantic properties associated to the male gender. This was true even when participants were presented with visual targets. Importantly, this male bias was persistent despite the fact that our pilot experiment on the facial images showed a slightly faster, although not statistically significant ( $p = 0.08$ ), tendency to process mixed pairs of faces. Our results therefore suggest that a strong male bias is indeed generated by the grammatical masculine form, and is not simply



an artifact of the experimental tasks employed in previous studies.

However, for both language groups, participants' responses to facial targets were also influenced by stereotypicality, with male stereotyped role nouns generating processing facilitation of following facial targets. In contrast, both response choices and positive response times indicated that facial targets following role nouns with a female stereotype were more difficult to process. We believe this to be indicative of an interference between the grammatically masculine form and the role noun's female stereotypicality. Namely, both sources of information compete, increasing processing difficulty. In contrast, an advantage was observed (i.e., a greater likelihood of allocating positive responses and an elicitation of faster response times) for targets following male stereotyped role nouns, which suggests that the congruency between the grammatically masculine gender and stereotypical gender facilitated participants' construction of mental representations.

Importantly, these main effects were further qualified by a consistent *stereotype* by *face pair* interaction for both the German and French group. This interaction indicated that participants' acceptance to face pairs changed as a function of the stereotypicality of the role noun preceding it. Male stereotyped role nouns triggered the greatest facilitation to accept male pairs of faces. These results support the idea that when reading a gender associated role noun such as *nurses* (*Krankenpfleger*<sub>German</sub>, *infirmiers*<sub>French</sub>), or *bosses* (*Arbeitgeber*<sub>German</sub>, *patrons*<sub>French</sub>) in a grammatical gender language, gender stereotypical information is immediately activated as part of the information associated with the role noun. As we did not embed our primes within sentences, our results suggest that this activation is made at the lexical access, with discursive text elements not needed to guide the activation of gender stereotypical information. Although we did find evidence that the masculine form was highly influential in guiding the representation toward a male-dominant representation as found in previous studies, we also documented that readers rely on immediate stereotypical information, even in the presence of a masculine grammatical form.

Our results, however, do not necessarily speak to whether, and to what extent, grammatical gender or stereotypical information has a greater influence over gender representations, as discussed in some discourse-based studies (Irmen, 2007). They mainly support the idea that both are activated at an early stage (i.e., lexical access), a claim that contrasts those of anaphor resolution studies that suggest an activation at later stages of comprehension (e.g., Irmen, 2007; Esaulova et al., 2013). The absence or weak indications of immediate stereotype effects in past studies could be attributed to several reasons. First of all, past research has frequently relied on verbal primes and verbal targets (e.g., Gyax and Gabriel, 2008; Gyax et al., 2012) to substantiate a persistent effect of the masculine form as specifically referring to men, with the effects of stereotype being only modest. The present study, however, demonstrated that the apparent lack of stereotype effects could be attributed to the tasks used to investigate these issues. We believe that by using facial images as targets, we went beyond simple language-on-language task

effects. Essentially the conceptual nature of stereotypes may have made them better candidates for non-verbal tasks which made it possible to delineate the true and noteworthy interaction between grammar and stereotypes when constructing a representation of gender. Another plausible argument for the absence of stereotype effects in past studies can be accredited to the nature of stereotype information, which dwindles rapidly as readers process discourse. Consequently, its effects did not clearly surface in previous studies on text comprehension. In the present study, the lexical-based paradigm may have allowed stereotype effects to surface before fading away, as they would have in a discursive context. Such a view may also support the reason for grammatical gender information to show a greater impact in most studies on the topic.

In terms of the impact of language shaping gender representations, the two language groups showed similar representation regularities in both their L1 and L2. This was rather unexpected given that we had anticipated the male bias to be reduced when participants processed the role nouns in German, due to its female-associated determiner. In fact, the German determiner did not elicit any substantial effects. Although there was a modest trend for mixed pairs of faces to be accepted more often when following female and neutral stereotyped role nouns (proportion of positive responses) when adding the determiner *die* for native German readers in L1, it did not lead to statistically significant effects. One could argue that when readers are faced with determining the grammatical gender of a noun, they will make use of available semantic (i.e., conceptual) and phonological information, which may result in processing facilitation (Schiller et al., 2003). In terms of our study, although both conceptual and masculine grammatical gender information competed to represent a probabilistic gender of the role noun, the association to the female gender of the German determiner did not substantially contribute in the representation process.

Although we cannot definitively refute the phenomenon, the male-attenuating effect in German documented by Rothermund (1998) appears to be at best superficial, at least in relation to the male-bias exerted by masculine forms. The fact that Garnham et al. (2012) found an effect of *sie* [they<sub>Female</sub>], was most likely due to the fact that they combined *die* and *sie*, both feminine equivalent, which offered a cumulative effect in deterring readers' attention from the role nouns' masculine form. In our German data, we observed only a main effect of *determiner*, whereby role nouns presented with a determiner facilitated responses to targets. These effects could be explained by the different rules associated to German. For instance, in French, although a noun must always be accompanied by a determiner even when a general statement is being made (e.g., *Infirmiers doivent s'occuper des personnes*. [Nurses need to care for people.] is grammatically incorrect: an article is always needed), in German, a noun can be presented both with and without a determiner (*Krankenpfleger müssen sich um Menschen kümmern*. [Nurses need to care for people.] vs. *Die Krankenpfleger müssen sich um Menschen kümmern*. [The nurses needed to care for people.]) which denote different meanings. The presence of *die* more clearly specifies that the role noun refers to a (particular) group of people, and not to

the general activity represented by the role noun, consequently facilitating subsequent associated targets. In this regard, our German group may have constructed different representations according to whether the role noun was presented with or without a determiner.

We thus believe that gender information associated to the determiner appears to be trivial, at least in comparison to the information associated to the gender inflection on the role noun. This gender inflection might be particularly relevant with person reference nouns, as they integrate conceptual gender as part of their lexical representation (Oakhill et al., 2005). In contrast, information linked to a function word, such as a determiner that connote less content and semantic information, would be less readily associated to any conceptual gender. Nonetheless, these results are in line with the numerous studies suggesting that the male bias exerted in grammatical gender languages is strong and appears to govern the comprehension processes. As such, our results substantiate the idea that language contributes in guiding mental representations. In our study, the grammatical masculine form contributed in shaping male-dominant representations across (more or less) all stereotypes, which is at odds with the idea that the masculine is the *unmarked* gender in grammatical gender languages. Although the impact of *die* in German was not observed, the effects of the masculine form of the role nouns lend support to the idea that grammatical markings may well direct (or bias) our attention to particular categories. The masculine form makes the male concept more accessible to readers. Note that this bias may not extend to less ambiguous cases such as *bigender* nouns in Italian, as investigated, for example, by Cacciari and Padovani (2007) or Siyanova-Chanturia et al. (2012).

Interestingly, we also observed a *task language by face pair* interaction surfacing in our German group's responses, suggesting that the male bias was more persistent in participants' L1 German than in their L2 French. This is crucial given that their dominant language exerted a greater male bias than their less fluent L2, despite having a better understanding and command of the language and the different interpretations of the masculine form in their L1. These results hint that the male bias stem from L1 for grammatical gender language readers. Such an account is in line with bilingual processing theories proposing that the languages of a bilingual are non-selectively activated even when only one language is being used for language comprehension processes (Dijkstra and van Heuven, 1998; de Groot et al., 2000).

Finally, we highlight that our linguistic-visual paradigm served as an effective approach to gauge the effects in question. The male bias and stereotype effects observed in our study were apparent in both the participants' L1 and in their less dominant L2. Importantly, despite the lack of stereotype effects observed in the presence of a strong masculine cue in past studies, our paradigm allowed us to observe stereotype effects. While some researchers argue that mixed-modal paradigms produce less priming effects (e.g., Alario

et al., 2000), our studies concurred with the conclusions made by Lemm et al. (2005) that they are still very efficient and powerful.

Our results suggest that thinking to speak or read in a grammatical gender language emphasizes gender associations, especially when these two are conceptually bound to each other. Although, our cognition of gender itself may not be fully influenced by grammatical gender, and this is an empirical question, our social cognition may well be, given that the concept of gender, especially that of male, is enhanced in grammatical gender language readers. These tendencies may then result in shifting or influencing our social perceptions of gender-stereotyped occupations, guiding readers to integrate a representation that is advantageous for men (Irmen and Köhncke, 1996; Braun et al., 2005).

## CONCLUSION

Using a linguistic-visual paradigm, the present study showed that readers automatically activate gender-associated information when reading gender stereotypical human referent role nouns. The activation of such information immediately takes place at a lexical level when readers encounter a role noun. Though morphological markings such as the default masculine form in French and German appear to be central when constructing mental representations of gender rather than superficial surface features, our study demonstrated that stereotype information also plays a role in influencing readers' mental representations. A stereotype effect was particularly apparent in the cumulative effects of stereotype and grammar when readers encounter male stereotyped role nouns. While past studies had not clearly found the effects of stereotype information in the presence of strong masculine effects (e.g., Gabriel and Gygas, 2008; Garnham et al., 2012; Gygas et al., 2012), the adaptation of a lexical and conceptual paradigm (with visual stimuli) was able to effectively gauge these effects. Future studies may want to further examine the possibilities of suppressing such male-dominant properties, though they appear to be relatively robust.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n° 237907.

## REFERENCES

- Alario, F.-X., Segui, J., and Ferrand, L. (2000). Semantic and associative priming in picture naming. *Q. J. Exp. Psychol. A* 53, 741–764. doi: 10.1080/713755907
- 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
- Banaji, M., and Hardin, C. (1996). Automatic stereotyping. *Psychol. Sci.* 7, 136–141. doi: 10.1111/j.1467-9280.1996.tb00346.x
- 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. M., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models using Eigen and S4* [R Package Version 1.1-7]. Available at: <http://CRAN.R-project.org/package=lme4>
- Boroditsky, L., Schmidt, L., and Phillips, W. (2003). “Sex, syntax, and semantics,” in *Language in Mind: Advances in the Study of Language and Cognition*, eds D. Gentner and S. Goldin-Meadow (Cambridge, MA: MIT Press), 61–80.
- Braun, F., Szeszny, S., and Stahlberg, D. (2005). Cognitive effects of masculine generics in German: an overview of empirical findings. *Communications* 30, 1–21. doi: 10.1515/comm.2005.30.1.1
- Brysbaert, M. (2007). “The Language-as-Fixed Effect Fallacy”: Some Simple SPSS Solutions to a Complex Problem. Egham: Royal Holloway, University of London.
- Bylund, E., and Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Biling. Lang. Cogn.* 14, 47–59. doi: 10.1017/S1366728910000180
- Cacciari, C., and Padovani, R. (2007). Further evidence of gender stereotype priming in language: semantic facilitation and inhibition in Italian role nouns. *Appl. Psycholinguist.* 28, 277–293. doi: 10.1017/S0142716407070142
- Carreiras, M., Garnham, A., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol.* 49A, 639–663. doi: 10.1080/027249896392531
- Cohen, J. D., MacWhinney, B., Flatt, M. R., and Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behav. Res. Methods Instrum. Comput.* 25, 257–271. doi: 10.3758/BF03204507
- Coleman, J. A. (1994). “Profiling the advanced language learner: the C-Test in British further and higher education,” in *Der C-Test: Theoretische Grundlagen und Praktische Anwendungen* [The C-test: Theoretical Foundations and Practical Applications], Vol. 2, ed. R. Grotjahn (Bochum: Brockmeyer), 217–237.
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- de Groot, A. M. B., Delmaar, P., and Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: support for non-selective access to bilingual memory. *Q. J. Exp. Psychol.* 53A, 397–428. doi: 10.1080/027249800390547
- 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. M. Jacobs (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 189–225.
- Eckes, T., and Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Lang. Test.* 23, 290–325. doi: 10.1191/0265532206lt330oa
- Esaulova, Y., Reali, C., and von Stockhausen, L. (2013). Influences of grammatical and stereotypical gender during reading: eye movements in pronominal and noun phrase anaphor resolution. *Lang. Cogn. Process.* 29, 781–803. doi: 10.1080/01690965.2013.794295
- Fausey, C. M., Long, B. L., Inamori, A., and Boroditsky, L. (2010). Constructing agency: the role of language. *Front. Psychol.* 1:162. doi: 10.3389/fpsyg.2010.00162
- Gabriel, U., and Gygax, P. (2008). Can societal language amendments change gender representation? The case of Norway. *Scand. J. Psychol.* 49, 451–457. doi: 10.1111/j.1467-9450.2008.00650.x
- Gabriel, U., Gygax, P., Sarrasin, O., Garnham, A., and Oakhill, J. (2008). Au pairs are rarely male: norms on the gender perception of role names across English, French, and German. *Behav. Res. Methods* 40, 206–212. doi: 10.3758/BRM.40.1.206
- Garnham, A., Gabriel, U., Sarrasin, O., Gygax, P., and Oakhill, J. (2012). Gender representation in different languages and grammatical marking on pronouns: when beauticians, musicians, and mechanics remain men. *Discourse Process.* 49, 481–500. doi: 10.1080/0163853X.2012.688184
- Gennari, S. P., Sloman, S. A., Malt, B. C., and Fitch, W. T. (2002). Motion events in language and cognition. *Cognition* 83, 49–79. doi: 10.1016/S0010-0277(01)00166-4
- Grotjahn, R., and Allner, B. (1996). “Der C-Test in der Sprachlichen Aufnahmeprüfung an Studienkollegs für ausländische Studierende an Universitäten in Nordrhein-Westfalen,” in *Der C-Test: Theoretische Grundlagen und Praktische Anwendungen*, Vol. 3, ed. R. Grotjahn (Bochum: Brockmeyer), 279–342.
- Grotjahn, R., Klein-Braley, C., and Raatz, U. (2002). “C-Tests: an overview,” in *University Language Testing and the C-Test*, eds J. A. Coleman, R. Grotjahn, and U. Raatz (Bochum: AKS-Verlag), 93–114.
- Gygax, P., and Gabriel, U. (2008). Can a group of musicians be composed of women? Generic interpretation of French masculine role names in the absence and presence of feminine forms. *Swiss J. Psychol.* 67, 143–151. doi: 10.1024/1421-0185.67.3.143
- Gygax, P., Gabriel, U., Lévy, A., Pool, E., Grivel, M., and Pedrazzini, E. (2012). The masculine form and its competing interpretations in French: when linking grammatically masculine role names to female referents is difficult. *J. Cogn. Psychol.* 24, 395–408. doi: 10.1080/20445911.2011.642858
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., and Garnham, A. (2008). Generically intended, but specifically interpreted: when beauticians, musicians, and mechanics are all men. *Lang. Cogn. Process.* 23, 464–485. doi: 10.1080/01690960701702035
- Hastings, A. J. (2002). “In defense of C-testing,” in *Der C-Test: Theoretische Grundlagen und Praktische Anwendungen*, Vol. 4, ed. R. Grotjahn (Bochum: AKS-Verlag), 11–29.
- Irmen, L. (2007). What’s in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *J. Psycholinguist. Res.* 36, 431–456. doi: 10.1007/s10936-007-9053-z
- Irmen, L., and Köhncke, A. (1996). Zur Psychologie des “generischen” Maskulinums [On the psychology of the “generic” masculine]. *Spr. Kogn.* 15, 152–166. doi: 10.1515/zfgl.33.2-3.212
- Kawakami, K., and Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Pers. Soc. Psychol. Bull.* 27, 212–225. doi: 10.1177/0146167201272007
- Kennison, S. M., and Trofe, J. L. (2003). Comprehending pronouns: a role for word-specific gender stereotype information. *J. Psycholinguist. Res.* 32, 355–378. doi: 10.1023/A:1023543602202
- Kreiner, H., Sturt, P., and Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: evidence from eye-movements. *J. Mem. Lang.* 58, 239–261. doi: 10.1016/j.jml.2007.09.003
- Lemm, K. M., Dabady, M., and Banaji, M. R. (2005). Gender picture priming: it works with denotative and connotative primes. *Soc. Cogn.* 23, 218–241. doi: 10.1521/soco.2005.23.3.218
- Oakhill, J., Garnham, A., and Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Mem. Cogn.* 33, 972–983. doi: 10.3758/BF03193206
- Papafragou, A., Massee, C., and Gleitman, L. (2002). Shake, rattle, “n” roll: the representation of motion in language and cognition. *Cognition* 84, 189–219. doi: 10.1016/S0010-0277(02)00046-X
- Pyykkönen, P., Hyönä, J., and van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: evidence from visual world eye tracking. *Exp. Psychol.* 57, 126–133. doi: 10.1027/1618-3169/a000002
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reali, C., Esaulova, Y., and von Stockhausen, L. (2015). Isolating stereotypical gender in a grammatical gender language: evidence from eye movements. *Appl. Psycholinguist.* 36, 977–1006. doi: 10.1017/S0142716414000010
- Reichert, M., Keller, U., and Martin, R. (2010). “The C-test, the TCF and the CEFR: a validation study,” in *Der C-Test: Beiträge Aus der Aktuellen Forschung. The*

- C-Test: Contributions from Current Research*, ed. R. Grotjahn (Frankfurt: Peter Lang), 205–231.
- Reynolds, D., Garnham, A., and Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *Q. J. Exp. Psychol.* 59, 886–903. doi: 10.1080/02724980543000088
- Rothermund, K. (1998). Automatische geschlechtsspezifische Assoziationen beim Lesen von Texten mit geschlechtseindeutigen und generisch maskulinen Text-Subjekten [Automatic gender-specific associations to texts containing gender-specific and masculine generic text subjects]. *Spr. Kogn.* 17, 183–198.
- Sato, S., Gygas, P. M., and Gabriel, U. (2013). Gender inferences: grammatical features and their impact on the representation of gender in bilinguals. *Biling. (Camb. Engl.)* 16, 792–807. doi: 10.1017/S1366728912000739
- Schiller, N. O., Münte, T. F., Horemans, I., and Jansma, B. M. (2003). The influence of semantic and phonological factors on syntactic decisions: an event-related brain potential study. *Psychophysiology* 40, 869–877. doi: 10.1111/1469-8986.00105
- Sigott, G., and Köberl, J. (1996). “Deletion patterns and C-test difficulty across languages,” in *Der C-Test: Theoretische Grundlagen und Praktische Anwendungen [The C-test: Theoretical Foundations and Practical Applications]*, ed. R. Grotjahn (Bochum: Brockmeyer), 159–172.
- Singular Inversions Inc (2004). *FaceGenModeller (Version Version 3.1.4)*. Available at: <http://www.FaceGen.com>
- Siyanova-Chanturia, A., Pesciarelli, F., and Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PLoS ONE* 7:e48712. doi: 10.1371/journal.pone.0048712
- Slobin, D. I. (1996). “From ‘thought and language’ to ‘thinking for speaking,’” in *Rethinking Linguistic Relativity*, eds J. J. Gumperz and S. C. Levinson (Cambridge: Cambridge University Press), 70–96.
- Stahlberg, D., Sczesny, S., and Braun, F. (2001). Name your favorite musician. *J. Lang. Soc. Psychol.* 20, 464–469. doi: 10.1177/0261927X01020004004
- Conflict of Interest Statement:** 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.
- The reviewer RHM and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.
- Copyright © 2016 Sato, Gygas and Gabriel. 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) or licensor 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.



# Effects of Word Width and Word Length on Optimal Character Size for Reading of Horizontally Scrolling Japanese Words

Wataru Teramoto<sup>1</sup>, Takuyuki Nakazaki<sup>2</sup>, Kaoru Sekiyama<sup>1</sup> and Shuji Mori<sup>3\*</sup>

<sup>1</sup> Division of Cognitive Psychology, Faculty of Letters, Kumamoto University, Kumamoto, Japan, <sup>2</sup> Mitsubishi Electric Corporation, Tokyo, Japan, <sup>3</sup> Department of Informatics, Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Ivlin Peev Stoitianov,  
Centre National de la Recherche  
Scientifique, France  
Katsuo Tamaoka,  
Nagoya University, Japan

### \*Correspondence:

Shuji Mori  
mori@inf.kyushu-u.ac.jp

### Specialty section:

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

Received: 04 June 2015

Accepted: 25 January 2016

Published: 16 February 2016

### Citation:

Teramoto W, Nakazaki T, Sekiyama K  
and Mori S (2016) Effects of Word  
Width and Word Length on Optimal  
Character Size for Reading  
of Horizontally Scrolling Japanese  
Words. *Front. Psychol.* 7:127.  
doi: 10.3389/fpsyg.2016.00127

The present study investigated, whether word width and length affect the optimal character size for reading of horizontally scrolling Japanese words, using reading speed as a measure. In Experiment 1, three Japanese words, each consisting of four Hiragana characters, sequentially scrolled on a display screen from right to left. Participants, all Japanese native speakers, were instructed to read the words aloud as accurately as possible, irrespective of their order within the sequence. To quantitatively measure their reading performance, we used rapid serial visual presentation paradigm, where the scrolling rate was increased until the participants began to make mistakes. Thus, the highest scrolling rate at which the participants' performance exceeded 88.9% correct rate was calculated for each character size (0.3°, 0.6°, 1.0°, and 3.0°) and scroll window size (5 or 10 character spaces). Results showed that the reading performance was highest in the range of 0.6° to 1.0°, irrespective of the scroll window size. Experiment 2 investigated whether the optimal character size observed in Experiment 1 was applicable for any word width and word length (i.e., the number of characters in a word). Results showed that reading speeds were slower for longer than shorter words and the word width of 3.6° was optimal among the word lengths tested (three, four, and six character words). Considering that character size varied depending on word width and word length in the present study, this means that the optimal character size can be changed by word width and word length in scrolling Japanese words.

**Keywords:** reading, Japanese, scrolling, character size, word perception

## INTRODUCTION

Reading is one of the most important functions facilitating intellectual activities in our daily life. Text for reading is provided statically by printed materials such as books and newspapers, and also dynamically by electronic devices such as the Times Square moving news display. The latter is, in general, termed scrolling (or drifting) text presentation, where a line of text drifts from right to left (or from bottom to top) along a single line. It is useful, especially when a large amount of information is to be displayed in a limited space such as train and building walls, websites, smartphones, and digital signage displays. Several previous studies have investigated the properties

of the visual mechanisms underlying reading performance with the scrolling text method, and have reported key parameters affecting reading text such as character size (Legge et al., 1985), contrast (Legge et al., 1987), spatial frequency (Legge et al., 1985), the number of characters simultaneously visible on the screen (Legge et al., 1985), color, and luminance (Legge and Rubin, 1986).

Character size is highly important among the factors determining the legibility of text. Legge et al. (1985) investigated the influence of character size on reading speeds with the scrolling text method and reported that the maximum reading speeds were achieved over a 10-fold range of character sizes from  $0.3^\circ$  to  $2.0^\circ$ . Reading speeds slowed down rapidly as character size decreased below  $0.2^\circ$ , and also slowed down gradually for characters larger than about  $2.0^\circ$ . In particular, they referred to the smallest character size below which reading speeds begin to decline rapidly, as critical print size (CPS). CPS is the limit of character size at which people can read text at optimal speed, but not an acuity limitation to identifying characters (Legge, 2007). Legge et al. (1985, 1987, 1989), Akutsu et al. (1991), Mansfield et al. (1996), and Chung et al. (1998) measured CPS in several studies and reported that CPS for normally sighted readers was almost constant (about  $0.2^\circ$ ) across different methods, although the overall reading speeds changed depending on the presentation method. Legge et al. (1989) compared the effect of character size on reading performance between scrolling and static text (i.e., printed text on paper) presentations. The results showed that CPS was  $0.25^\circ$  for both methods, but at  $0.25^\circ$  the difference in reading speed between them was a maximum of 250 words/min. Chung et al. (1998) used the rapid serial visual presentation (RSVP) paradigm (Rubin and Turano, 1992) to avoid the potential influence of oculomotor control and demonstrated that the average CPS across six subjects was  $0.17^\circ$ . Note that, while an average reading speed was 1171 words/min for RSVP text (Rubin and Turano, 1992), it was 250–300 words/min for scrolling text (Legge et al., 1985). These findings suggest that character size plays a crucial role in reading performance.

The present study further investigated the effect of character size on reading of scrolling text, focusing on word length (the number of characters in a word) and word width (visual angle of a word). Most of the previous studies used sentences as test stimuli. Thus, words of various lengths (and word width) were included in the test stimuli. From the viewpoint of eye movements in reading, it is well known that word length and word width are crucial factors for controlling saccade eye movements during reading of static text: fixation duration is longer for longer than shorter words (Just and Carpenter, 1980; Rayner et al., 1996; Calvo and Meseguer, 2002; Kliegl et al., 2004) and the number of fixations is larger for longer than shorter words (Rayner and McConkie, 1976; Rayner et al., 1996; Kliegl et al., 2004). Morrison and Rayner (1981) measured the saccade distance during reading of static sentences, while manipulating viewing distances. The results showed that the saccade distance was constant if measured in number of character spaces (5.3–5.7 character spaces, while the corresponding saccade size was  $2.0^\circ$  to  $3.8^\circ$ ). It should be noted that the fixation duration increased with increasing viewing

distance (i.e., a decrease of visual angle). McDonald (2006) investigated the genuine effect of word length on eye movements in reading, by presenting all words at the same visual angle within a sentence. They revealed that the fixation duration and the number of fixations increased with an increase in the number of characters in the word. Hautala et al. (2011) confirmed the findings of McDonald (2006) with their stimuli of a proportional font (Arial) and further showed that word width influenced other aspects of saccade eye movements, such as the fixation location in words and the probability of skipping a word. An increase in fixation duration and the number of fixations would normally be indicative of a decrease in reading speed. Thus, these findings suggest a possibility that word length and word width impact on reading speed. In fact, Legge et al. (1997, 2001) investigated the effect of word length on reading performance, using the RSVP paradigm, in order to clarify the relationship between reading speed and visual span, i.e., the number of characters that can be recognized without eye movements (O'Regan, 1990). Carver (1976) also pointed out the effect of word length on reading speed in terms of text difficulty (i.e., easier text generally includes shorter words, so reading speed is higher). These studies revealed that word length had an impact on reading performance of static text, especially at low contrast and in peripheral vision (Legge et al., 1997, 2001). However, no study has thus far investigated how word length and word width influence reading performance of scrolling text and, specifically, whether the character size effect previously observed in scrolling text is applicable for any word length and word width.

In Experiment 1, we investigated the effect of character size on reading speed in scrolling text presentation, using Japanese 4-character words written in Japanese Hiragana, which are similar in usage to the letters of English, but different in phonologic complexity (i.e., a syllable consists of a single letter in Hiragana, but one to several letters in English). In Experiment 2, we investigated whether the effect of character size observed in Experiment 1 was applicable for any word width and word length. The results showed that word width and word length substantially influenced the optimal character size in reading.

## EXPERIMENT 1

### Method

#### Participants

There were 10 participants (aged 20–28; all males). All the participants had normal or corrected-to-normal vision and were unaware of the purpose of the experiment. All were native Japanese speakers. This study followed the tenets of the Declaration of Helsinki, and the protocol was approved by the ethics committee of Faculty of Information Science and Electrical Engineering, Kyushu University. Written informed consent was obtained from all participants (including those of Experiment 2) prior to participation.

#### Experimental Design

A two-by-four within-participant factorial design was used: scroll window size (5 and 10 characters) and character size ( $0.3^\circ$ ,  $0.6^\circ$ ,

1.0°, and 3.0°). The scroll window size was defined as character spaces in which a character moves on the screen from right to left. This is different from “scroll distance,” which is defined as character spaces in which a word moves, corresponding to the sum of the window size and word length. Since both scroll window size and scroll distance were based on character space, their visual angles varied depending on character size. Previous studies found that reading speeds increased with an increase in window size up to 4 (Legge et al., 1985) or 4.7 (Fine and Peli, 1996a) and were no further increase or decrease for the larger character sizes. By introducing, two types of scroll window size beyond five character spaces in the present study, we intended to confirm if the optimal character size observed in one scroll window size condition could be generalized to that in the other scroll window size. Note that character size, character space and character width were defined based on the center-to-center spacing of the adjacent characters throughout this paper<sup>1</sup>.

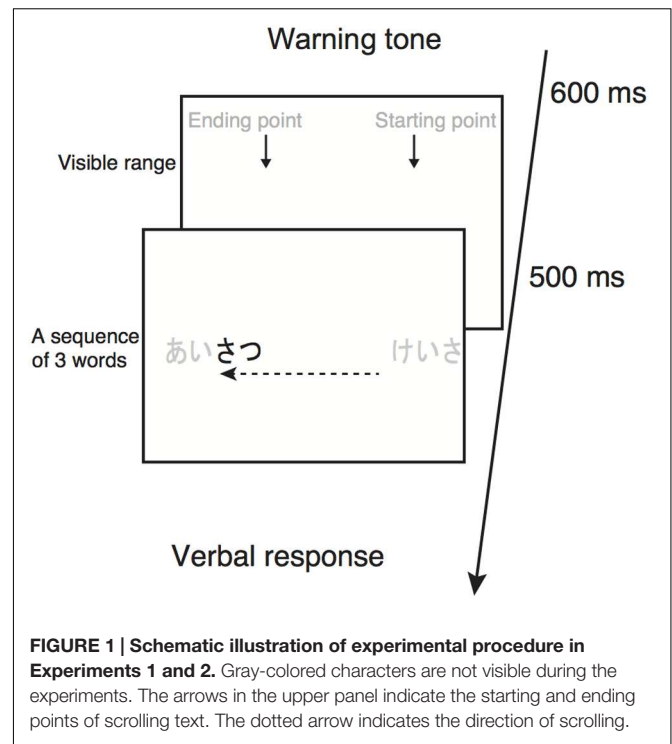
### Apparatus and Stimuli

Visual stimuli were generated using a visual stimulus generator (ViSaGe; Cambridge Research Systems, UK), attached to an IBM-compatible personal computer (Dell PRECISION 390) and displayed on a  $\gamma$ -corrected 19-in. computer monitor (EIZO, Flex Scan T760) with a refresh rate of 100 Hz and a resolution of 1,024 × 768. All visual stimuli were presented in the center of the monitor.

All the words used were selected within the range of high word familiarity, i.e., over 5.4 (maximum: 7) in the Japanese familiarity-controlled word lists (Amano and Kondo, 2003). In each trial, test words were chosen randomly from a pool of 2,223 words in order to avoid the effect of context on reading performance (Fine and Peli, 1996b). All the words presented contained four Hiragana characters and four morae (phonological sound units that determine syllable weights) because the average word length in Legge et al. (1985) was 4.1 characters so our results would be comparable to their results using 4-character words. Moreover, we found in the preliminary experiment that 4-character words were appropriate, because they were not easily identified upon sight of the first one or two characters, but shorter character words were very easy to identify. Also, reading performance for words containing more than six characters was degraded drastically. The words were rendered in MS Gothic Japanese, a fixed-width font, and were presented as black characters (0.39 cd/m<sup>2</sup>) on a white background of 68.62 cd/m<sup>2</sup>.

### Procedure

Experiments were conducted with binocular viewing in a dark room. Immediately after the participant's button press, a warning tone was presented for 600 ms, followed by a 500-ms presentation of two arrows indicating the starting and ending points of scrolling text (Figure 1). Soon after the disappearance of the



**FIGURE 1 | Schematic illustration of experimental procedure in Experiments 1 and 2.** Gray-colored characters are not visible during the experiments. The arrows in the upper panel indicate the starting and ending points of scrolling text. The dotted arrow indicates the direction of scrolling.

arrows, three Japanese words were displayed consecutively<sup>2</sup>, one word at a time. Although the participants were able to look anywhere they wanted on the display before the presentation of the words because of no fixation point on the display, they normally looked around the starting point of scrolling. The participants were instructed to read the words aloud as accurately as possible, irrespective of the word order within the sequence. They were allowed to complete their verbalization during the presentation of words as well as after the words disappeared from the display. There was no time constraint for their responding. The experimenter assessed their responses and entered the number of correct responses into the computer. Immediately after that, a warning tone was presented for 600 ms and the next trial began.

The scrolling speed was manipulated by changing the exposure duration of a word on the monitor—the time taken from the appearance of the initial character until the disappearance of the fourth character. The initial exposure duration of each word was determined according to the performance of each participant in the preliminary trials so that it was sufficiently long (i.e., slow in scrolling speed) for each participant to report all the words. In subsequent trials, the exposure duration varied from trial to trial according to a staircase method (Cornsweet, 1962). The exposure duration was decreased when eight or nine words out of nine words (three words × the last three trials) were correctly reported (88.9%

<sup>1</sup>As a reference, we measured inter-character white space by using a maximum width character. The proportion of lateral white spacing of the character-to-character size is about 0.03°. That is, in the case where a character with a size of 1.0° is used, each of its left and right white spaces is 0.03° and the width of character *per se* is 0.94°.

<sup>2</sup>We did not control the overlapping of phonemes and syllables in three consecutive words in each trial. The sequence of presented words was randomly selected from a large word list in each participant and the same word did not appear until the list was completed. Thus, we think that the overlap effect on the whole reading performance would be very small, if any.

correct). The exposure duration was decreased by 80 ms for the first decrement, 40 ms for the second decrement, and 10 ms for subsequent decrements. Conversely, the exposure duration was increased by 10 ms when the number of words correctly reported was less than eight out of nine words. The staircase was terminated after eight reversals of the exposure duration sequence. The threshold of the exposure duration was calculated as the average of all but the first two reversals. We confirmed in the preliminary experiment that the last six reversals were enough to produce the stable threshold. Each participant performed one staircase per each character size per each scroll window size. The average number of trials they performed was  $17.9 \pm 3.4$  (standard deviation). Reading speed in words per minutes (Legge et al., 1985; Carver, 1990) was also calculated, using the threshold (in ms), according to the following equation:

$$\text{Reading speed (words/min)} = \frac{60000}{\text{threshold exposure duration}}$$

Thus, the reading speed was defined as the number of words readable for 1 min with 88.9% accuracy under a given condition. The optimal character size was defined as the character size with which reading performance reached the maximum speed.

To present each character clearly on the monitor, two viewing distances were adopted. The viewing distance was 172 cm for the character sizes of  $0.3^\circ$  and  $0.6^\circ$  and 57 cm for the character sizes of  $1.0^\circ$  and  $3.0^\circ$ . Different viewing distances were tested in different blocks. Half of the participants performed the 172-cm viewing distance block first and the other half performed the 57-cm viewing distance block first. The order of conditions tested within a block was randomized. It took about 80 min for each participant to complete all conditions.

## Results and Discussion

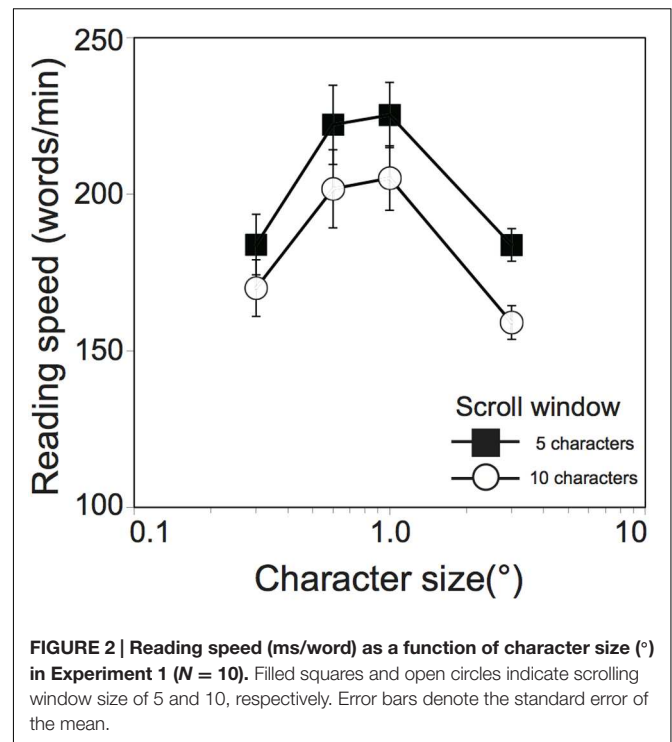
Threshold exposure durations obtained in Experiment 1 are shown in **Table 1**. Averages of calculated reading speeds across participants are shown as a function of character size in **Figure 2**. The reading speed increased with an increase in character size, reached a maximum with  $1.0^\circ$  of character size for both the 5-character ( $225 \pm 10.4$  words/min) and 10-character ( $205 \pm 10.3$  words/min) scroll window size conditions, and decreased for the larger character size. Since Shapiro–Wilk tests revealed that the data were normally distributed in all variables ( $W \geq 0.881$ ,  $p \geq 0.132$ ), the reading speeds were analyzed in a two-factor analysis of variance (ANOVA) with within-participants factors<sup>3</sup>

<sup>3</sup>In this and the subsequent ANOVAs, the significant levels were computed using Huynh–Feldt corrected degrees of freedom for lack of sphericity in repeated measures.

**TABLE 1 | Average threshold exposure durations in Experiment 1 (ms).**

Scroll window size (char.)	Character size ( $^\circ$ )			
	0.3	0.6	1.0	3
5	334 (52.6)	278 (48.4)	272 (39.6)	329 (29.3)
10	362 (60.5)	307 (55.4)	299 (45.3)	382 (43.2)

Numbers in parenthesis represent standard deviations.



**FIGURE 2 | Reading speed (ms/word) as a function of character size ( $^\circ$ ) in Experiment 1 ( $N = 10$ ).** Filled squares and open circles indicate scrolling window size of 5 and 10, respectively. Error bars denote the standard error of the mean.

(scroll window size and character size). A main effect of scroll window size was significant [ $F(1,9) = 33.01$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.103$ ]: the reading was faster for the 5-character than for the 10-character scroll window size conditions. A main effect of character size was also significant [ $F(3,27) = 27.15$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.458$ ], but there was no significant interaction between them [ $F(3,27) = 1.97$ ,  $p = 0.142$ ,  $\eta_G^2 = 0.004$ ]. A Tukey’s *post hoc* test for the character size factor ( $\alpha < 0.05$ ,  $MSe = 386.58$ ) revealed that reading speed was significantly faster for the  $0.6^\circ$  and  $1.0^\circ$  than the  $0.3^\circ$  and  $3.0^\circ$  conditions. The optimal character size observed in this experiment are almost consistent with those of previous studies using scrolling English words (e.g., Legge et al., 1985; Akutsu et al., 1991), indicating that the effect of character size on reading performance can be generalized beyond differences in tested words (i.e., Japanese Hiragana words) and the experimental procedure. It should be noted that the maximum reading speeds (225 words/min and 205 words/min for the 5-character and 10-character scroll window size conditions, respectively) were slower than those of previous studies (about 300 words/min; Legge et al., 1985; Akutsu et al., 1991). This might be from numerous factors such as language (Japanese vs. English), context (random words vs. continuous sentences), experimental procedure (one word at a time vs. sentences), and so on.

Although the optimal character size was consistent between the scroll distance conditions, the overall reading speed was larger for the 5-character than for the 10-character scroll window size conditions. However, it should be noted that reading speeds in the present definition are derived only from the threshold exposure time. In the case where the threshold exposure time is identical between 5-character and 10-character scroll



window size conditions, reading speed (words/min) would be the same while scrolling speed ( $^{\circ}$ /s) would be higher for the 10-character than the 5-character conditions. This is because the exposure time measure disregards the difference in scroll distance. In order to appreciate the effect of scroll window size on reading performance, we computed a new reading performance measure, “scrolling speed at threshold,” according to the following equation:

$$\text{Scrolling speed at threshold (character spaces/s)} = \frac{(\text{scroll window size} + 4)}{(\text{threshold exposure duration}/1000)}$$

where ‘4’ in the numerator of the right-hand side is the word length, i.e., four characters (see “Experimental Design”). Average scrolling speeds at thresholds across the participants are shown as a function of character size in **Figure 3**. The scrolling speed at threshold increased with an increase in character size, reached a maximum with  $1.0^{\circ}$  of character size for both the 5-character ( $33.8 \pm 4.7$  character spaces/s) and 10-character ( $47.9 \pm 7.2$  character spaces/s) scroll window size conditions, and decreased for the larger character size. Overall, the scrolling speeds were larger for the 10-character than 5-character scroll window size conditions. Since Shapiro–Wilk tests revealed that the data were normally distributed in all variables ( $W \geq 0.881$ ,  $p \geq 0.132$ ), a two-way ANOVA with within-participants factors (scroll window size and character size) was performed. The analysis revealed that both main effects and an interaction were significant [scroll window size,  $F(1,9) = 153.55$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.539$ ; character size,  $F(3,27) = 28.30$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.306$ ; scroll

window size  $\times$  character size,  $F(3,27) = 11.56$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.024$ ]. The analysis of the interaction revealed that maximum scrolling speeds were significantly larger for the 10-character than 5-character scroll window size conditions across all character size [ $0.3$ ,  $F(1,36) = 114.54$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.713$ ;  $0.6$ ,  $F(1,36) = 148.26$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.763$ ;  $1.0$ ,  $F(1,36) = 155.42$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.771$ ;  $3.0$ ,  $F(1,36) = 71.30$ ,  $p < 0.001$ ,  $\eta_G^2 = 0.607$ ]. This new analysis suggests that the slower reading speed in the 10- than 5-character scroll window size conditions could be also explained by the difference in scrolling speeds. Thus, to properly evaluate the effect of the scroll window size on reading performance, a more sophisticated word presentation method is necessary. Of importance in the present study is the fact that the reading performance was highest in the range of  $0.6^{\circ}$  to  $1.0^{\circ}$ , irrespective of the scroll window size tested.

## EXPERIMENT 2

Experiment 1 demonstrated that reading performance was highest in the range of  $0.6^{\circ}$  to  $1.0^{\circ}$ , irrespective of the scroll window size. The effect of scroll window size was also observed. Since Experiment 1 used only 4-character words, the word width was changed depending on the character size so that there could be an interaction between character size and word width. Thus, Experiment 2 investigated whether the optimal character size observed in Experiment 1 was applicable for any word width and word length. If character size is the only crucial factor for reading, reading performance should be highest in the range of  $0.6^{\circ}$  to  $1.0^{\circ}$  and degraded out of that range, irrespective of word width and word length.

## Method

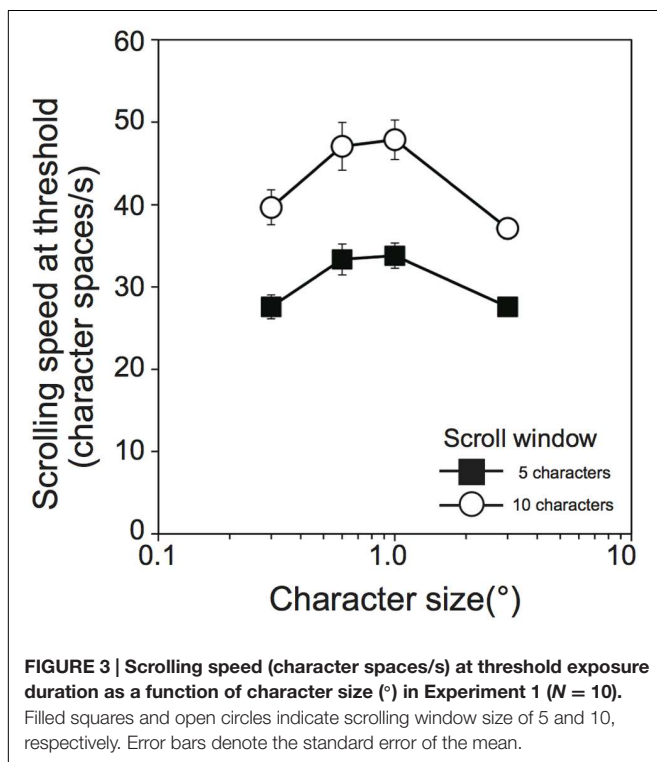
### Participants

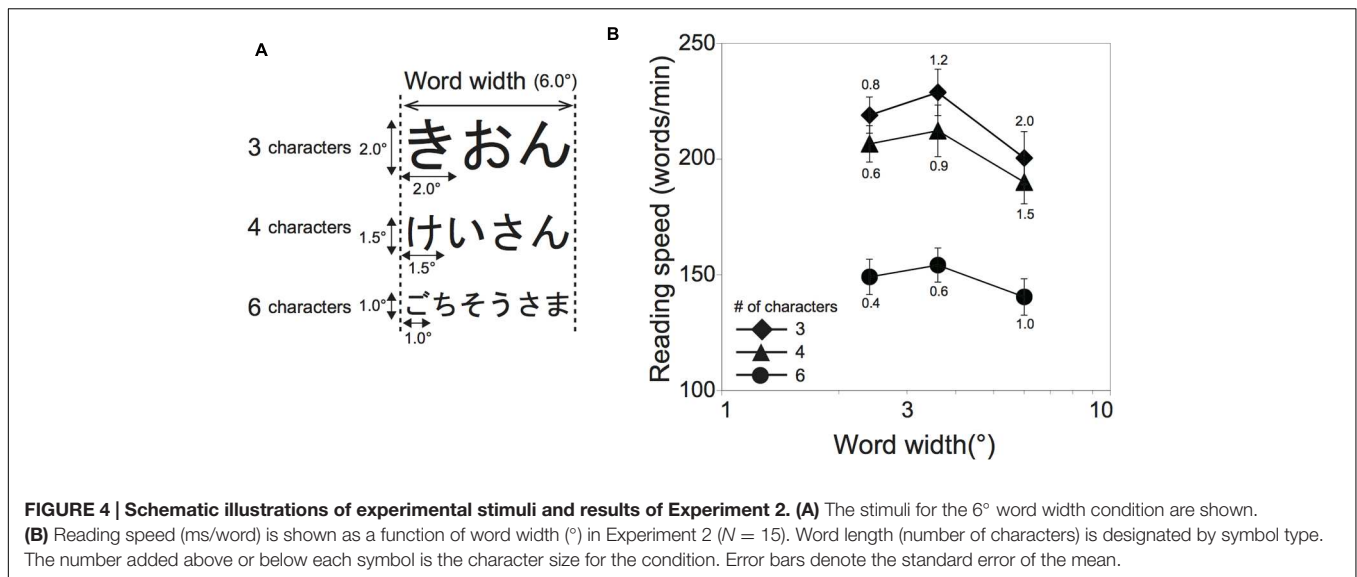
There were 15 participants (aged 19–25; all males). All the participants had normal or corrected-to-normal vision and were unaware of the purpose of the experiment. All were native Japanese speakers.

### Experimental Design, Stimuli, and Procedure

A three-by-three within-participant factorial design was used: word length (the number of characters in a word; three, four, and six characters) and word width ( $2.4^{\circ}$ ,  $3.6^{\circ}$ , and  $6.0^{\circ}$ ). The character size was adjusted according to the word length and the word width (see **Figure 4A**). For example, the character sizes for three character words for  $2.4^{\circ}$ ,  $3.6^{\circ}$ , and  $6.0^{\circ}$  of word width were  $0.8^{\circ}$ ,  $1.2^{\circ}$ , and  $2.0^{\circ}$ , respectively, while those for four character words were  $0.6^{\circ}$ ,  $0.9^{\circ}$ , and  $1.5^{\circ}$ , respectively, and those for six character words were  $0.4^{\circ}$ ,  $0.6^{\circ}$ , and  $1.0^{\circ}$  respectively. If character size is the only crucial factor for reading, reading performance should be highest for  $2.4^{\circ}$  of word width in the 3-character word condition, for  $2.4^{\circ}$  and  $3.6^{\circ}$  of word width in the 4-character condition and for  $3.6^{\circ}$  and  $6.0^{\circ}$  of word width in the 6-character condition.

Three and four character words were selected within the range of high word familiarity, i.e., over 5.4 (maximum: 7) in





the Japanese familiarity-controlled word lists, while six character words were selected within the range of word familiarity over 5.0 in the list because of fewer samples. In each trial, test words were chosen randomly from a pool of 1,518 words for three character words, 1,598 words for four character words, and 937 words for six character words. All the words were presented in Hiragana only. The words were rendered in MS Gothic Japanese, a fixed-width font, and were presented as black characters (0.86 cd/m<sup>2</sup>) on a white background of 105.13 cd/m<sup>2</sup>. The scroll window size was 10 character spaces across conditions.

To present each character clearly on the monitor, two viewing window sizes were adopted. The viewing window size was 172 cm for the character size under 0.8° and 57 cm for character size more than 0.8°. Different combinations of word length and word width were tested in different sessions. The order of conditions tested was randomized. Thus, the participants performed nine experimental sessions.

### Results and Discussion

Threshold exposure durations obtained in Experiment 2 are shown in **Table 2**. Averages of calculated reading speeds across participants are shown as a function of word width in **Figure 4B**. Each line indicates the number of characters in a word. The number added above or below each symbol was the character

size for the condition. Shorter word lengths resulted in higher reading speeds for each word width. There was a peak of reading speed at the word width of 3.6°, irrespective of the number of characters in a word. Since Shapiro–Wilk tests revealed that the data were not normally distributed in a variable ( $W = 0.887, p = 0.011$ ), we applied logarithmic transformation to the whole data (the normality test results after transformation:  $W \geq 0.888, p \geq 0.063$ ). A two-factor ANOVA with within-participants factors (word length and word width) was performed on the transformed data. There were significant main effects of word length [ $F(2,28) = 84.91, p < 0.001, \eta_G^2 = 0.484$ ] and word width [ $F(2,28) = 26.35, p < 0.001, \eta_G^2 = 0.074$ ] and no significant interaction between them [ $F(4,56) = 0.64, p = 0.637, \eta_G^2 = 0.002$ ]. Tukey’s *post hoc* tests for the word length factor ( $\alpha < 0.05, MSe = 0.004$ ) revealed that the reading speed was significantly different among these word length conditions: the 6-character condition obtained slower reading speed than the others. Notably, the *post hoc* tests for the word width factor ( $\alpha < 0.05, MSe = 0.001$ ) revealed that the reading speed for the 6.0° word width condition was slower than those for the other conditions. Experiment 1 showed the highest reading performance for the 0.6° and 1.0° character size conditions. However, the data of Experiment 2 demonstrated that reading performance became worse for the 1.0° than 0.4° character size condition if words extended to 6.0° wide and consisted of six characters. These results indicate that word width and word length have an impact on the optimal character size at least in the range of character size tested.

**TABLE 2 | Average threshold exposure durations in Experiment 2 (ms).**

Word length (char.)	Word width (°)		
	2.4	3.6	6.0
3	279 (38.2)	269 (41.1)	311 (57.4)
4	296 (43.4)	293 (52.8)	326 (57.5)
6	416 (75.2)	402 (72.8)	445 (88.0)

Numbers in parenthesis represent standard deviations.

### GENERAL DISCUSSION

The present study investigated the influence of word width and word length on the optimal character size for reading of horizontally scrolling Japanese words, by using reading speed as a measure. The results of Experiment 1 demonstrated that reading performance was highest at the character sizes of 0.6°

to 1.0°, irrespective of the scroll window size (5-character and 10-character spaces). Experiment 2 investigated, how word width and word length influence the effect of character size observed in Experiment 1. The results showed that reading speed increased with a decrease in word length and that the reading performance became worse even for the character size of 1.0°, with which the highest reading performance was obtained in Experiment 1 if words extended to 6.0° wide and consisted of six characters. These findings suggest that both word width and word length can influence the optimal character size in reading scrolling text.

The effect of character size on reading has been intensively investigated. A common finding is that the maximum reading speed is obtained over a wide range of character sizes (0.2°–2.0° in Legge et al., 1985), and declines sharply below CPS and gradually for larger character sizes. The results of Experiment 1 are consistent with the previous studies: reading speed was nearly constant for 0.6° and 1.0° of character size, and declined for 0.3° and 3.0°. More important results of the present study are that the effect of character size can be dependent on word width and word length as shown in Experiment 2. Most of the previous studies used sentences as test stimuli so that words of various widths and lengths were included. Therefore, it was unclear whether these parameters can influence the effect of character size on reading. Experiment 2 showed that, for the condition in which words extended 6.0° wide and consisted of six characters, the reading performance became worse even for the 1.0° character size condition, which had yielded the maximum reading speed in Experiment 1. The effects of word length and word width on reading have already been reported in studies investigating eye-movements during reading. Longer words receive a longer fixation duration and a greater number of fixations than shorter words (Rayner and McConkie, 1976; Just and Carpenter, 1980; Rayner et al., 1996; Calvo and Meseguer, 2002; Kliegl et al., 2004). These findings imply that word width and word length can have significant impacts on reading speed as well, because the fixation duration and the number of fixations are closely related to reading speed. The present study provides evidence that word width and word length can influence not only eye movements in reading but also reading speed.

The effects of word width and word length observed in the present study demand an explanation that deals with the visual processing of character strings rather than individual characters (i.e., character acuity, contrast sensitivity, and so on). We speculate that concepts of visual span (Legge et al., 2001) and uncrowded span (Pelli et al., 2007) can explain our findings. Legge et al. (2001) developed a psychophysical method (trigram method) to measure visual span, the number of adjacent characters that can be read without moving eyes. Specifically, they presented strings of three characters at several positions left and right of the fixation, so briefly that no eye movements could occur. The participants were asked to report all three characters in left-to-right order. Character recognition accuracies (percent correct rates) were plotted as a function of distance of the fixation. The size of the visual span was defined as the distance from the fixation for which character recognition accuracy exceeded 80% correct. Legge et al. (2001, 2007) and Yu et al. (2007, 2010) found a high correlation between reading speed and the size of the visual

span. This indicates that visual span could be linked to reading speed. However, it was not clear what determined the size of the visual span. Pelli et al. (2007) introduced a concept of “crowding” of character recognition (Bouma, 1970, 1973) into reading and demonstrated that crowding imposes the major limitation of the size of the visual span (see also Levi et al., 2007). In other words, they suggest that it is not character size but spacing between characters that limits the size of the visual span, and, as a result, reading speed. The critical character spacing, the minimum center-to-center spacing between target and flankers with which the target character can be identified at a threshold level, is proportional to eccentricity and extends 0.1° in the normal fovea (Bouma, 1970). Recognition of characters is prerequisite for recognition of words. Thus, when more than one character fall within a critical character spacing, not only character recognition but also word recognition is spoiled. The important point is that the critical character spacing is not uniform across a word. Words extend horizontally (and sometimes vertically) so they have several critical character spacings. Thus, there is a high possibility that more eccentric characters exceed the critical character spacing, especially within longer and wider words. This would be likely to explain our results in Experiment 2, where the reading performance for longer and wider words became worse even for the 1.0° character size condition, which had yielded maximum reading speed in Experiment 1.

## CONCLUSION

The present study investigated the influence of word width and word length on the optimal character size for reading of horizontally scrolling words. Results showed that the reading performance became worse even for the character size of 1.0°, with which the highest reading performance was obtained with four character words, when the presented words extended to 6.0° wide and consisted of six characters. These findings suggest that both word width and word length can influence the optimal character size in reading scrolling Japanese text at least.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

We are grateful to two reviewers for their valuable and insightful comments and suggestions about the data analysis and interpretation for early versions of the manuscript. This research is based on the Master's thesis of the second author, submitted to Kyushu University. We would like to thank Syouta Yoshimura for his technical support. This research was supported by JSPS Grant-in-Aids for Scientific Research (A) (No. 25240023) and (B) (No. 21330169) to SM, and JSPS Grant-in-Aid for Challenging Exploratory Research (No. 21653078) to KS.

## REFERENCES

- Akutsu, H., Legge, G. E., Ross, J. A., and Schuebel, K. J. (1991). Psychophysics of reading—X. Effects of age-related changes in vision. *J. Gerontol.* 46, 325–331. doi: 10.1093/geronj/46.6.P325
- Amano, N., and Kondo, T. (2003). *NTT Database Series Nihongo-no Goitokusei (Lexical Properties of Japanese) Part 1 [CD]*. Tokyo: Sanseido.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature* 226, 177–178. doi: 10.1038/226177a0
- Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Res.* 13, 767–782. doi: 10.1016/0042-6989(73)90041-2
- Calvo, M. G., and Meseguer, E. (2002). Eye movements and processing stages in reading: relative contributions of visual, lexical and contextual factors. *Span. J. Psychol.* 5, 66–77. doi: 10.1017/S113874160005849
- Carver, R. P. (1976). Word length, prose difficulty and reading rate. *J. Read. Behav.* 8, 193–204.
- Carver, R. P. (1990). *Reading Rate: A Review of Research and Theory*. San Diego, CA: Academic Press.
- Chung, S. T., Mansfield, J. S., and Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Res.* 38, 2949–2962. doi: 10.1016/S0042-6989(98)00072-8
- Cornsweet, T. N. (1962). The staircase method in psychophysics. *Am. J. Psychol.* 75, 485–491. doi: 10.2307/1419876
- Fine, E. M., and Peli, E. (1996a). Visually impaired observers require a larger window than normally sighted observers to read from a scroll display. *J. Am. Optom. Assoc.* 67, 390–396.
- Fine, E. M., and Peli, E. (1996b). The role of context in reading with central field loss. *Optom. Vis. Sci.* 73, 533–539. doi: 10.1097/00006324-199608000-00003
- Hautala, J., Hyönä, J., and Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Res.* 51, 1719–1727. doi: 10.1016/j.visres.2011.05.015
- Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354. doi: 10.1037/0033-295X.87.4.329
- Kliegl, R., Grabner, E., Rolfs, M., and Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *Eur. J. Cogn. Psychol.* 16, 262–284. doi: 10.1080/09541440340000213
- Legge, G. E. (2007). *Psychophysics of Reading in Normal and Low Vision*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Legge, G. E., Ahn, S. J., Klitz, T. S., and Luebker, A. (1997). Psychophysics of reading—XVI. The visual span in normal and low vision. *Vision Res.* 37, 1999–2010. doi: 10.1016/S0042-6989(97)00017-5
- Legge, G. E., Cheung, S. H., Yu, D., Chung, S. T., Lee, H. W., and Owens, D. P. (2007). The case for the visual span as a sensory bottleneck in reading. *J. Vis.* 7, 9.1–9.15. doi: 10.1167/7.2.9
- Legge, G. E., Mansfield, J. S., and Chung, S. T. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Res.* 41, 725–743. doi: 10.1016/S0042-6989(00)0295-9
- Legge, G. E., Pelli, D. G., Rubin, G. S., and Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Res.* 25, 239–252. doi: 10.1016/0042-6989(85)90117-8
- Legge, G. E., Ross, J. A., Luebker, A., and LaMay, J. M. (1989). Psychophysics of reading. VIII. The Minnesota low-vision reading test. *Optom. Vis. Sci.* 66, 843–853. doi: 10.1097/00006324-198912000-00008
- Legge, G. E., and Rubin, G. S. (1986). Psychophysics of reading. IV. Wavelength effects in normal and low vision. *J. Opt. Soc. Am. A* 3, 40–51. doi: 10.1364/JOSAA.3.000040
- Legge, G. E., Rubin, G. S., and Luebker, A. (1987). Psychophysics of reading—V. The role of contrast in normal vision. *Vision Res.* 27, 1165–1177. doi: 10.1016/0042-6989(87)90028-9
- Levi, D. M., Song, S., and Pelli, D. G. (2007). Amblyopic reading is crowded. *J. Vis.* 7, 21.1–21.17. doi: 10.1167/7.2.21
- Mansfield, J. S., Legge, G. E., and Bane, M. C. (1996). Psychophysics of reading. XV. Font effects in normal and low vision. *Invest. Ophthalmol. Vis. Sci.* 37, 1492–1501.
- McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Vis. Cogn.* 13, 89–98. doi: 10.1080/13506280500143367
- Morrison, R. E., and Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. *Percept. Psychophys.* 30, 395–396. doi: 10.3758/BF03206156
- O'Regan, J. K. (1990). "Eye movements and reading," in *Eye Movements and Their Role in Visual and Cognitive Processes*, ed. E. Kowler (New York, NY: Elsevier), 395–453.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., and Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *J. Vis.* 7, 20.1–20.36. doi: 10.1167/7.2.20
- Rayner, K., and McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Res.* 16, 829–837. doi: 10.1016/0042-6989(76)90143-7
- Rayner, K., Sereno, S. C., and Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *J. Exp. Psychol. Hum.* 22, 1188–1200. doi: 10.1037/0096-1523.22.5.1188
- Rubin, G. S., and Turano, K. (1992). Reading without saccadic eye movements. *Vision Res.* 32, 895–902. doi: 10.1016/0042-6989(92)90032-E
- Yu, D., Cheung, S. H., Legge, G. E., and Chung, S. T. (2007). Effect of letter spacing on visual span and reading speed. *J. Vis.* 7, 2.1–2.10. doi: 10.1167/7.2.2
- Yu, D., Park, H., Gerold, D., and Legge, G. E. (2010). Comparing reading speed for horizontal and vertical English text. *J. Vis.* 10, 21.1–21.17. doi: 10.1167/10.2.21

**Conflict of Interest Statement:** 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 © 2016 Teramoto, Nakazaki, Sekiyama and Mori. 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) or licensor 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.



# Electrophysiological Correlates of Second-Language Syntactic Processes Are Related to Native and Second Language Distance Regardless of Age of Acquisition

Begoña Díaz<sup>1\*</sup>, Kepa Erdocia<sup>2</sup>, Robert F. de Menezes<sup>1</sup>, Jutta L. Mueller<sup>3,4</sup>,  
Núria Sebastián-Gallés<sup>1</sup> and Itziar Laka<sup>2</sup>

<sup>1</sup> Center for Brain and Cognition, Department of Technology, Universitat Pompeu Fabra, Barcelona, Spain, <sup>2</sup> Department of Linguistics and Basque Studies, Faculty of Arts, University of the Basque Country, Vitoria-Gasteiz, Spain, <sup>3</sup> Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, <sup>4</sup> Psycho- and Neurolinguistics Group, Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Maria Garraffa,  
Heriot-Watt University, UK  
Roha Mariam Thomas,  
Oklahoma State University, USA

### \*Correspondence:

Begoña Díaz  
begona.diaz@upf.edu

### Specialty section:

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

**Received:** 01 July 2015

**Accepted:** 25 January 2016

**Published:** 12 February 2016

### Citation:

Díaz B, Erdocia K, de Menezes RF,  
Mueller JL, Sebastián-Gallés N and  
Laka I (2016) Electrophysiological  
Correlates of Second-Language  
Syntactic Processes Are Related to  
Native and Second Language  
Distance Regardless of Age of  
Acquisition. *Front. Psychol.* 7:133.  
doi: 10.3389/fpsyg.2016.00133

In the present study, we investigate how early and late L2 learners process L2 grammatical traits that are either present or absent in their native language (L1). Thirteen early (AoA = 4 years old) and 13 late (AoA = 18 years old) Spanish learners of Basque performed a grammatical judgment task on auditory Basque sentences while their event-related brain potentials (ERPs) were recorded. The sentences contained violations of a syntactic property specific to participants' L2, i.e., ergative case, or violations of a syntactic property present in both of the participants' languages, i.e., verb agreement. Two forms of verb agreement were tested: subject agreement, found in participants' L1 and L2, and object agreement, present only in participants' L2. Behaviorally, early bilinguals were more accurate in the judgment task than late L2 learners. Early bilinguals showed native-like ERPs for verb agreement, which differed from the late learners' ERP pattern. Nonetheless, approximation to native-likeness was greater for the subject-verb agreement processing, the type of verb-agreement present in participants' L1, compared to object-verb agreement, the type of verb-agreement present only in participants' L2. For the ergative argument alignment, unique to L2, the two non-native groups showed similar ERP patterns which did not correspond to the natives' ERP pattern. We conclude that non-native syntactic processing approximates native processing for early L2 acquisition and high proficiency levels when the syntactic property is common to the L1 and L2. However, syntactic traits that are not present in the L1 do not rely on native-like processing, despite early AoA and high proficiency.

**Keywords:** bilingualism, morphosyntax, event-related potentials, P600, age of acquisition, language distance

## INTRODUCTION

In a growing global world, learning a second language (L2) has become a socioeconomic need and, consequently, is a mandatory subject in most countries in the world (around 81% of the 119 countries analyzed in (UNESCO World Report: Investing in Cultural Diversity Intercultural Dialogue., 2009). There are three main elements for L2 learners to conquer: phonology (perception and pronunciation), semantics (the meaning of words), and syntax (the structure of the language). Syntax (along with phonology) is an especially difficult aspect for L2 learners to master, while semantics is more easily overcome (Johnson and Newport, 1989; Van Hell and Tokowicz, 2010). One main question in psycholinguistics is whether an L2 is processed through the same neural mechanisms engaged in L1 processing. Event-related potential (ERP) studies have shown clear components indexing L1 syntactic processes and have revealed that approximation to native-like processing of an L2 syntax depends on three main factors: the age of acquisition (AoA), the proficiency achieved, and the similarity between one's first language (L1) and the target L2 (for a review of studies in relation to AoA, proficiency and language distance factors, see Kotz, 2009; Caffarra et al., 2015). Here we aim to investigate how the syntactic similarity between the L1 and L2 influences L2 syntactic processing. We address this question by comparing two groups of non-native listeners that represent the two endpoints of the AoA-proficiency continuum: a typical group of late learners with average proficiency and another group of early learners with native-like proficiency.

Early acquisition favors L2 learning and native-like processing, as shown by the landmark study of Weber-Fox and Neville (1996). They studied Chinese (L1)—English (L2) bilinguals divided into groups according to their L2-AoA: 1–3, 4–6, 7–10, 11–13, and those above the age of 16. Participants performed a visual grammatical judgment task involving semantic violations, as well as three different syntactic violations (phrase structure, specificity constraint, and subadjacency constraint). Semantic violations elicited the typical N400 (a negativity around 400 ms) for semantic processing in natives and all non-native groups. For syntactic violations, the ERPs were native-like when the L2 was learned before the age of 11 and not native-like when learned after the age of 11. Natives and early L2 learners (AoA < 11 years old) displayed a left-anterior negativity (LAN) followed by a P600. The LAN is an ERP component elicited by grammatical violations between 300 and 500 ms. The P600 is a posterior positivity starting around 600 ms that indexes syntactic reanalysis and repair. For all those who learned after the age of 11, the LAN was bilaterally distributed. In addition, the P600 was delayed for the 11–13 age group and no P600 was reported for the >16 age group. They concluded that maturational constraints for L2 learning exist, in line with the Critical Period Hypothesis (Lenneberg, 1967), and that native-like neural mechanisms are used for L2 syntactic processing if learned before puberty. After this study, many others have reported non-native ERP patterns for late L2 learners (Hahne and Friederici, 1999; Hahne, 2001; Mueller et al., 2005, 2007; Ojima et al., 2005).

A criticism for the AoA view is that late L2 learners usually achieve lower L2 proficiency than early L2 learners, making it difficult to tell which of the two factors, AoA or proficiency, is the origin of the non-native brain responses (Friederici et al., 2002; Rossi et al., 2006; Kotz et al., 2008). Rossi et al. (2006) recorded the ERP responses of late learners (AoA > 10 years old). Half of the participants were German natives learning Italian, and the other half of the participants were Italian natives learning German. Learners were divided into high and low proficiency groups and submitted to an auditory grammatical judgment task in the L2. High proficiency groups displayed ERP patterns typically found during native language processing: an early left anterior negativity (ELAN) between 100 and 200 ms that is claimed to index syntactic parsing, followed by the syntactic reanalysis and repair P600 response for word category violations, and a LAN-P600 response for subject-verb agreement violations. Low proficiency groups showed a very distinct ERP pattern: they only showed a delayed and reduced P600 in response to both syntactic violations. These results suggested that when late L2 learners attain high proficiency in the L2, they process the L2 in a native-like fashion.

Conversely, other studies have shown non-native ERP patterns during L2 processing despite high proficiency (Ojima et al., 2005; Chen et al., 2007; Dowens et al., 2011; Pakulak and Neville, 2011; Zawiszewski et al., 2011; Erdocia et al., 2014). For instance, Pakulak and Neville (2011) compared the ERP responses to English phrase structure violations of English listeners and late German learners of English (AoA > 10 years old) during an auditory grammatical judgment task. Importantly, the two groups were matched in their English proficiency as measured with a standardized, norm-referenced test and the grammatical judgment task. Despite being equally accurate in the task as the natives, the late L2 learners displayed a non-native ERP response, namely a broadly distributed P600, whereas the natives displayed an anterior negativity followed by a P600. This finding refuted the claim that high proficiency results in native-likeness even when the L2 is learned late in life. Chen et al. (2007) and Ojima et al. (2005) studied the processing of English subject-verb agreement in late learners (AoA = 12 years old) who did not had such a grammatical trait in their native language. Chen et al. (2007) studied Chinese listeners with a visual grammatical judgment task and Ojima et al. (2005) studied Japanese listeners with a visual comprehension task. The processing of verb agreement violations elicited a LAN and a P600 in natives in both studies. However, non-natives did not show any P600 despite being highly proficiency in English, and a LAN was only observed in Ojima et al. (2005). The authors argued that the absence of a similar syntactic structure in the native language of the learners may be responsible for the distinct neural responses between learners and natives. Indeed, studies reporting native-like processing in late L2 learners with high proficiency studied the processing of syntactic features present in both of the participants' languages. For instance, Rossi et al. (2006) studied word category and agreement violations, which are syntactic traits of both of the participants' languages (i.e., German and Italian). Likewise, Kotz et al. (2008) studied phrase structure violations, a syntactic trait

also present in both of the participants' languages (i.e., Spanish and English).

Several studies have investigated the role that L1–L2 similarity plays in L2 syntactic processing. These studies consistently show native-likeness (or an approximation) in highly proficient, late bilinguals for syntactic traits that are common to the L1 and the L2 but non-native ERPs for grammatical traits that are unique to the L2 (Tokowicz and MacWhinney, 2005; Osterhout et al., 2006; Sabourin and Stowe, 2008; Dowens et al., 2010; Foucart and Frenck-Mestre, 2011). This results however were not replicated in Tokowicz and MacWhinney (2005) and Aleman Bañon et al. (2014), who reported native-like ERPs for Spanish gender agreement violations in late English learners of Spanish. However, grammatical gender constitutes a highly heterogeneous class across languages encompassing phonological, morphological, lexical, and syntactic phenomena (Corbett, 1991, 1994). Grammatical gender, thus, contrasts with other linguistic traits like word order or ergativity in that no systematic typological correlations are associated to the presence or absence of gender in a given language (Greenberg, 1978). Crucial evidence supporting the major impact that language similarity has on L2 processing comes from studies showing non-native processing even in the most favorable scenario: when native-like proficiency is attained in a frequently used language learned early in life. Zawiszewski et al. (2011) studied the ERP responses elicited by the processing of syntactic traits that were L1–L2 similar or unique to L2 in early Spanish (L1)—Basque (L2) bilinguals (L2 AoA > 3 years old) during a written grammatical judgment task. For the L1–L2 similar condition, i.e., object-verb agreement, a native-like pattern was found (all participants showed a N400–P600 pattern) while a non-native pattern was found for the conditions unique to the L2. For the unique head parameter (subject-object-verb (SOV) and SVO word order) condition, both groups displayed similar P600 effects, but natives showed a left parietal negativity between 300 and 500 ms, while non-natives showed a frontally distributed negativity in the same time window followed by a broad negativity between 500 and 600 ms. For the unique ergative condition, all participants showed a broadly distributed negativity, but only the natives displayed a P600. Erdocia et al. (2014) also reported non-native processing of syntactic traits unique to L2 with the same type of population studied by Zawiszewski et al. (2011), i.e., highly proficient, early Spanish (L1)—Basque (L2) bilinguals. With a written sentence comprehension task, they studied the processing of sentences following the canonical word order of the L2 (SOV) or a non-canonical order (OSV), both of which differed from the canonical L1 word order (SVO). The comparison of the O and S elements in the second position of the sentence elicited a P600 in early bilinguals but a left temporal–posterior negativity between 400 and 550 ms in natives. Thus, Zawiszewski et al. (2011) and Erdocia et al. (2014) showed that, despite early acquisition and high proficiency, the processing of divergent L1–L2 traits involved non-native processing mechanisms.

Overall, findings on L2 syntactic processing suggest that native-like brain mechanisms are only engaged in processing L2 syntactic traits that are present in the L1, while AoA and proficiency seem to influence the efficiency of the native-like brain mechanisms (as measured by the amplitude and latency

of the ERPs). Nevertheless, the impact that AoA and proficiency have in the processing of syntactic traits unique to L2 has not been systematically studied. The present study aims to investigate the role of AoA and proficiency in the processing of syntactic traits that diverge in L1 and L2. To the best of our knowledge, this is the first study comparing early and late L2 learners in their processing of syntactic traits that differ in the similarity between L1 and L2 employing the same experimental procedures.

In the present study, 25 early Spanish (L1)—Basque (L2) bilinguals (AoA = 4 years of age) and 25 late L2 learners (AoA > 16 years of age) participated in an auditory grammatical judgment task on Basque (L2) sentences while their EEG signal was recorded. Spanish and Basque are typologically very different and provide the opportunity to compare grammatical traits that are either present or absent in participants' L1 (Spanish). Spanish is a Romance language, while Basque is a language isolate (De Rijk, 2007). Both Spanish and Basque have verb agreement, agreeing in person and number between a verb and its arguments. However, Spanish only agrees with the subject (*1a*), whereas Basque has multipersonal agreement, meaning the verb must agree with subject and object concurrently (*1b*).

- (1) *a.* Tú me has visto. / *b.* Zuk ni ikusi nauzu.  
 You me have-2.s seen / You-erg me- $\emptyset$ (abs) seen 1.s-have-2.s  
 “You have seen me.”

Spanish and Basque also possess different argument alignment types. Spanish is a nominative-accusative language (like English or German), while Basque has an ergative-absolutive alignment (like Hindi or Georgian). They also have distinct canonical word orders (Spanish is an SVO language, whereas Basque is an SOV language). Basque also differs from Spanish due to its morphological case marking, i.e., the overt marking of the core arguments of the sentence with specific morphemes (see *1b*), a trait not present in Spanish (see *1a*).

In the present auditory grammatical judgment task, half of the sentences heard were grammatically correct, while the other half had either a subject-verb agreement violation, an object-verb agreement violation, or an ergative case violation (see **Table 1**). Both subject- and object-verb agreement violations presented a mismatch in number between a plural subject or object, respectively, and a corresponding singular agreement marker in the auxiliary verb. Following Zawiszewski et al. (2011), subject- and object-verb agreement are considered as similar in L1 and L2. Note, however, that Spanish, participants' L1, possesses only subject-verb agreement. For the ergative case violation, two noun phrases marked for ergative case were presented in one sentence. Ergative case is unique to L2. In a previous study (Díaz et al., 2011), native Basque listeners were presented with exactly the same task and materials<sup>1</sup>. Native Basque listeners showed a P600 component, an index of syntactic repair and reanalysis processes,

<sup>1</sup>The present study does not present a direct comparison of the natives and non-natives data sets because of differences in the number of participants for each group (24 native listeners were tested in the previous study vs. 13 participants in each non-native group) and differences in the analysis procedures (natives' brain responses were analyzed with EProbe (ANT, The Netherlands) and no eye correction was applied, whereas non-natives' responses were analyzed with BrainVision Analyzer 2.0 (Brain Products GmbH, Munich, Germany) and eye movements were corrected.

**TABLE 1 | Experimental stimulus examples.**

## 1. Grammatical Sentence (40 items)

Mikel-en **arreb-ek egunkari-a** saski-a-n ekarri **d-u-te** kiosko-tik

Mikel-[gen.] **sister-the-[erg.pl.] newspaper-the-[abs.sg.]** basket-the-in brought **it-root-they** kiosk-from

'Mikel's sisters have brought the newspaper in a basket from the kiosk'

## 2. Subject-Verb Agreement Violation (40 items)

Mikel-en **arreb-ek** egunkari-a saski-a-n ekarri **d-\*u** kiosko-tik

Mikel-[gen.] **sister-the-[erg.pl.] newspaper-the-[abs.sg.]** basket -the-in brought **it-root-\*it** kiosk-from

## 3. Ergative Case Violation (40 items)

Mikel-en **arreb-ek egunkari-\*ek** saski-a-n ekarri d-u-te kiosko-tik

Mikel-[gen.] **sister-the-[erg.pl.] newspaper-the-\*[erg.pl.]** basket -the-in brought it-root-they kiosk-from

## 4. Grammatical Object-Verb Agreement (40 items)

Mikel-en arreb-ek **egunkari-a-k** saski-a-n ekarri **dit-u-zte** kiosko-tik

Mikel-[gen.] sister-the-[erg.pl.] **newspaper-the-[abs.pl.]** basket -the-in brought **them-root-they** kiosk-from

'Mikel's sisters have brought the newspapers in a basket from the kiosk'

## 5. Object-Verb Agreement Violation (40 items)

Mikel-en arreb-ek **egunkari-a-k** saski-a-n ekarri **\*d-u-te** kiosko-tik

Mikel-[gen.] sister-the-[erg.pl.] **newspaper-the-[abs.pl.]** basket -the-in brought **\*it-root-they** kiosk-from

*Bold words represent the critical word for each violation condition from which onset epochs were established.*

in response to these three syntactic violations. In addition, the object-verb agreement violation elicited an early posterior negativity between 150 and 300 ms. The finding of a P600 for subject-verb agreement, object-verb agreement, and ergative case violations is in line with previous studies with native listeners across several languages, such as English, Spanish, Basque, and Hindi (Coulson et al., 1998a,b; Frisch and Schlesewsky, 2005; Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; Zawiszewski et al., 2011). In addition, in previous studies with Basque and Hindi native listeners, an N400 was found for object-verb agreement and ergative case violations (Nevins et al., 2007; Zawiszewski and Friederici, 2009; Zawiszewski et al., 2011). This N400 effect was interpreted as an index of costs in computing thematic relationships. However, no N400 was found for any of the violations for Basque natives in Díaz et al. (2011) with the same materials used in the present study. Regarding object-verb agreement, the different agreement feature tested in Díaz et al. (2011) and previous studies (Zawiszewski and Friederici, 2009; Zawiszewski et al., 2011), i.e., number vs. person, respectively, could be the reason for the different ERP pattern. It has been suggested that person plays a more salient role in agreement computations than number, based on the finding of larger P600s for person compared to number agreement violations (Nevins et al., 2007; Mancini et al., 2011; Zawiszewski et al., in press). In line with these previous studies, the violation of the person feature in agreement, as compared to number violations, could lead to higher costs in thematic assignments (Díaz et al., 2011). Regarding ergative case violations, the incorrect sentences in Díaz et al. (2011) always had a correct ergative marked noun phrase, whereas in the two previous studies, the case violation sentences did not include a correct ergative NP (Nevins et al., 2007; Zawiszewski et al., 2011). The lack of a correct ergative NP could

cause thematic difficulty in assigning the agency of the ergative argument, as reflected by the N400. Thus, the differences in the specific characteristics of the experimental materials between Díaz et al. (2011) and previous studies (Nevins et al., 2007; Zawiszewski and Friederici, 2009; Zawiszewski et al., 2011) could be the cause for the distinct ERP patterns observed.

In the present study, we expect native-like ERP responses (i.e., a P600 and an additional early negativity for object-verb agreement as in Díaz et al., 2011) in highly proficient, early bilinguals for L2 grammatical traits present in participants' L1 (verb-agreement conditions). In contrast, we expect non-native ERP responses in the same bilingual group for the grammatical trait unique to L2 (ergative case condition). Additionally, the comparison of the results for the two types of verb agreement violations, subject and object, will allow us to investigate whether L2 verb agreement relations are similarly processed. In less proficient, late L2 learners, we expect non-native effects for all conditions. The critical question is whether AoA and proficiency also have an impact on the processing of the unique L2 trait. The unique L2 trait is expected to elicit non-native ERP patterns in both groups of L2 learners. Nevertheless, we expect that highly proficient, early bilinguals will show a different ERP pattern from less proficient, late learners. The differences between the groups for the processing of the unique L2 trait would reveal what the correlates of L2 mastery in non-native ERPs are.

## MATERIALS AND METHODS

### Participants

Fifty healthy adult participants took part in the experiment. All participants were born and grew up in the Basque

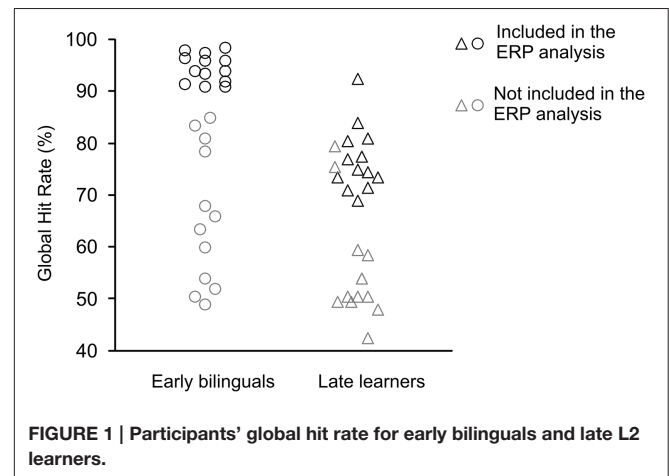


Country where both Spanish and Basque are spoken. A Basque adaptation of the language history questionnaire from Weber-Fox and Neville (1996) was administered to all participants. This questionnaire assessed the relative use of Spanish and Basque during childhood, adolescence, and at the time of evaluation. In addition, participants rated their own Spanish and Basque proficiency. No participant reported having had auditory, language or neurological problems. All participants were right-handed as assessed with the Edinburgh Handedness Inventory (Oldfield, 1971). All participants signed the corresponding consent form and were paid for their participation. The experiment was approved by the local ethical committee of the University of the Basque Country and followed the American Psychological Association standards in accordance with the Declaration of Helsinki (World Medical Association, 2013).

The group of early bilinguals tested was composed of 25 university students, Spanish-Basque bilinguals (15 female, mean age: 22.68 years old, range 19–30 years old). For all participants, Spanish was the family language from birth to the time of testing. Thus, prior to attending school, participants had just sporadic (if any) contact with Basque. Participants were continuously exposed to Basque from the age of 3 or 4 when starting mandatory bilingual school. Participants were recruited from the University of the Basque Country.

The group of late L2 learners tested was made up of 25 Spanish-Basque bilinguals (17 female, mean age: 26.71 years old, range 19–36 years old). All late L2 learners were Spanish monolinguals who were, at the time and for 2 years prior, attending classroom-based Basque instruction. They were all enrolled in their fourth semester of Basque lessons, thus on their way toward completing a B2 level (Common European Framework of Reference for Languages). They started Basque instruction at a mean age of 24.70 ( $SD = 4.49$ ). Participants were recruited from several *euskaltegi* (official schools dedicated exclusively to teaching Basque to adults) in the Vitoria-Gasteiz area (Basque Country). As all *euskaltegi* centers follow the same curriculum, late L2 learners can be assumed to have the same knowledge of Basque and did not have virtually any contact with Basque language prior to attending Basque lessons. Despite both Spanish and Basque being official languages in the Basque Country, there is a part of the Basque population that has no contact with the Basque language. According to a sociolinguistic survey published by the Basque government (V Encuesta Sociolingüística: 2011 (2003)), 27% of Basque citizens older than 16 are fluent in both Spanish and Basque, 14.7% can understand it but not speak it, and the remaining 58.3% are Spanish monolinguals. All late L2 learners except one had received, or were receiving at the time, higher education (university, college, or apprenticeship studies).

Not all participants were included in the ERP analysis. Late L2 learners' accuracy in the grammatical judgment task varied greatly (see **Figure 1**, Results) from chance levels (accuracy between 40 and 60%) to relatively good proficiency (accuracy  $\geq 69\%$ ). Only those late L2 learners with a global accuracy of 69% or above in the task were included in the ERP data analysis. Fifteen late L2 learners had the minimum accuracy required,



**FIGURE 1 |** Participants' global hit rate for early bilinguals and late L2 learners.

but due to artifacts in the EEG signal, the data from two of those late L2 learners were excluded from all the analyses. To match the number of participants in each group, 13 early bilinguals with the highest accuracy were included in the ERP analysis. **Table 2** shows participants' characteristics and self-reported relative language use through life span for the sample of participants included in the ERP analysis.

## Stimuli

**Table 1** displays examples of the grammatical and ungrammatical sentences used in the present study. The words in the sentences were all present in the late L2 learners' Basque textbooks. In addition, the grammatical structures tested (verb agreement and ergative case) were early topics in the Basque lessons. Forty grammatical Basque sentences were created. The subject-verb agreement and ergative case violations were derived from the grammatical sentences. Subject-verb agreement violation sentences were created by a mismatch in number between plural subjects and singular verb agreement. Ergative case violation sentences had two arguments with the ergative case. A second grammatical set of sentences was created for comparisons with the object-verb agreement violations. The second set of grammatical sentences was identical to the first set of grammatical sentences except that a plural object agreed with the verb. We created the ungrammatical sentences by changing the grammatical auxiliary verb to a singular object-verb agreement auxiliary.

The experimental sentences were presented with 80 grammatical filler sentences. The critical words were never the last word of the sentence to avoid wrap-up effects. Sentences were digitally recorded at 16-bits by a native, female Basque speaker in a soundproof booth. The sentences across conditions were similar in mean amplitude and length [amplitude:  $F_{(6, 234)} = 1.36, p > 0.05$ ; length:  $F_{(6, 234)} < 1$ ].

## Procedure

The ERP recordings were conducted in a soundproof room at the Psycholinguistics Laboratory (University of the Basque Country in Vitoria-Gasteiz). Participants sat in front of a computer

**TABLE 2 | Group characteristics and self-reported relative use of Spanish and Basque during life span, ranging from 1 (Basque only) to 7 (Spanish only), and self-reported proficiency, ranging from 1 (perfect) to 4 (poor).**

	Early bilinguals ( <i>n</i> = 13)	Late L2 learners ( <i>n</i> = 13)
Age*	23.23 (3.03)	26.76 (5.16)
AoA of Basque*	3.23 (0.43)	24.76 (5.16)
Sex (females)	8	10
<b>RELATIVE LANGUAGE USE</b>		
<b>Before school</b>		
Home	6.30 (1.25)	6.92 (0.27)
<b>Primary school</b>		
Home*	6.23 (0.83)	7.00 (0.00)
School*	3.07 (1.80)	6.38 (0.76)
Others*	5.15 (1.46)	6.84 (0.37)
<b>Secondary school</b>		
Home*	5.92 (1.18)	6.92 (0.27)
School*	2.76 (1.58)	6.38 (0.50)
Others*	4.38 (1.60)	6.76 (0.59)
<b>At time of testing</b>		
Home*	5.30 (1.49)	6.84 (0.37)
University/Work*	2.46 (1.39)	6.46 (0.66)
Others*	4.07 (1.03)	5.61 (1.32)
<b>SELF-RATED PROFICIENCY</b>		
Global Proficiency*	1.00 (0.00)	2.53 (0.66)
Comprehension*	1.00 (0.00)	2.38 (0.76)
Speaking*	1.23 (0.43)	3.15 (0.89)
Writing*	1.38 (0.50)	2.84 (0.89)

Standard deviations are in parentheses. \*Significant differences between early bilinguals and late L2 learners (two-sample *t*-test comparisons).

monitor in a comfortable armchair. They received written instructions in their L2 (Basque). Participants were instructed to perform a delayed grammatical judgment task (programmed with EXPE6; Pallier et al., 1997). Participants were asked to listen to the sentences and respond whether it was incorrect or correct for each sentence. Responses were given by pushing one of the buttons held in each hand. The correspondence between correct and incorrect responses and hands was counterbalanced across participants. Participants were told about the importance of being still during the ERP recordings. In addition, they were asked to avoid eye movements (including blinking) during the trials. Participants were free to blink between trials when a resting message appeared on the screen. Participants first performed a training of eight practice trials with feedback. For the experiment, participants performed 320 trials. The trials followed a pseudo-random order that did not allow the presentation of more than three successive trials of the same condition. Trials started with a fixation point (an asterisk) for 500 ms, which was then followed by an auditory sentence. Sentences were played

binaurally through headphones (Sennheiser HD 435 Manhattan). The asterisk remained on the screen during the full sentence and for 1500 ms after sentence offset. The asterisk was then replaced by a written message that prompted participants to respond. There was no time limit for participants' response. The next trial started 1500 ms after participants' response. A message was presented on the screen during this inter-trial interval that informed participants they could blink freely.

## Electrophysiological Recording

The EEG was recorded with the BrainVision 2.0 Analyzer Software package and a BrainAmp amplifier (Brain Products). The EEG signal was recorded from the scalp using tin electrodes mounted in an electro-cap (Electro-Cap International). Electrodes were located at 58 standard positions (Fp2, Fpz, Fp1, F4A, F3A, F2, Fz, F1, F4, F3, F6, F5, F7, F8, C2A, CZA, C1A, C4A, C3A, C6A, C5A, C2, Cz, C1, C4, C3, C6, C5, C2P, C1P, C4P, C3P, T4, T3, T6, T5, T4L, T3L, P2, P1, P6, P5, CB2, CB1, P2P, PZA, P1P, TCP2, TCP1, P4, Pz, P3, P4P, PZP, P3P, O1, Oz, O2). Electrodes attached the outer canthus and to the infra-orbital ridge of the right eye measured eye movements. The EEG recording was referenced online to the right mastoid and re-referenced offline to linked mastoids. Electrode impedances were kept below 5 k $\Omega$ . The EEG signal was filtered online with a band-pass between 0.01 and 50 Hz and digitized at a sampling rate of 500 Hz.

## Data Analysis

### Behavioral Data

Hit rates were calculated for each participant and condition. **Figure 1** displays the global proficiency of the participants, which was calculated by averaging the hit rates for each participant across all conditions. Many late L2 listeners showed poor accuracy levels in the grammatical judgment task. To select those late learners with sufficient accuracy in the grammatical judgment task we compared the hit rates of each late learner for the five conditions against chance level (50%) by means of one-sample *t*-tests. Only those that performed above chance were included in the ERP analysis.

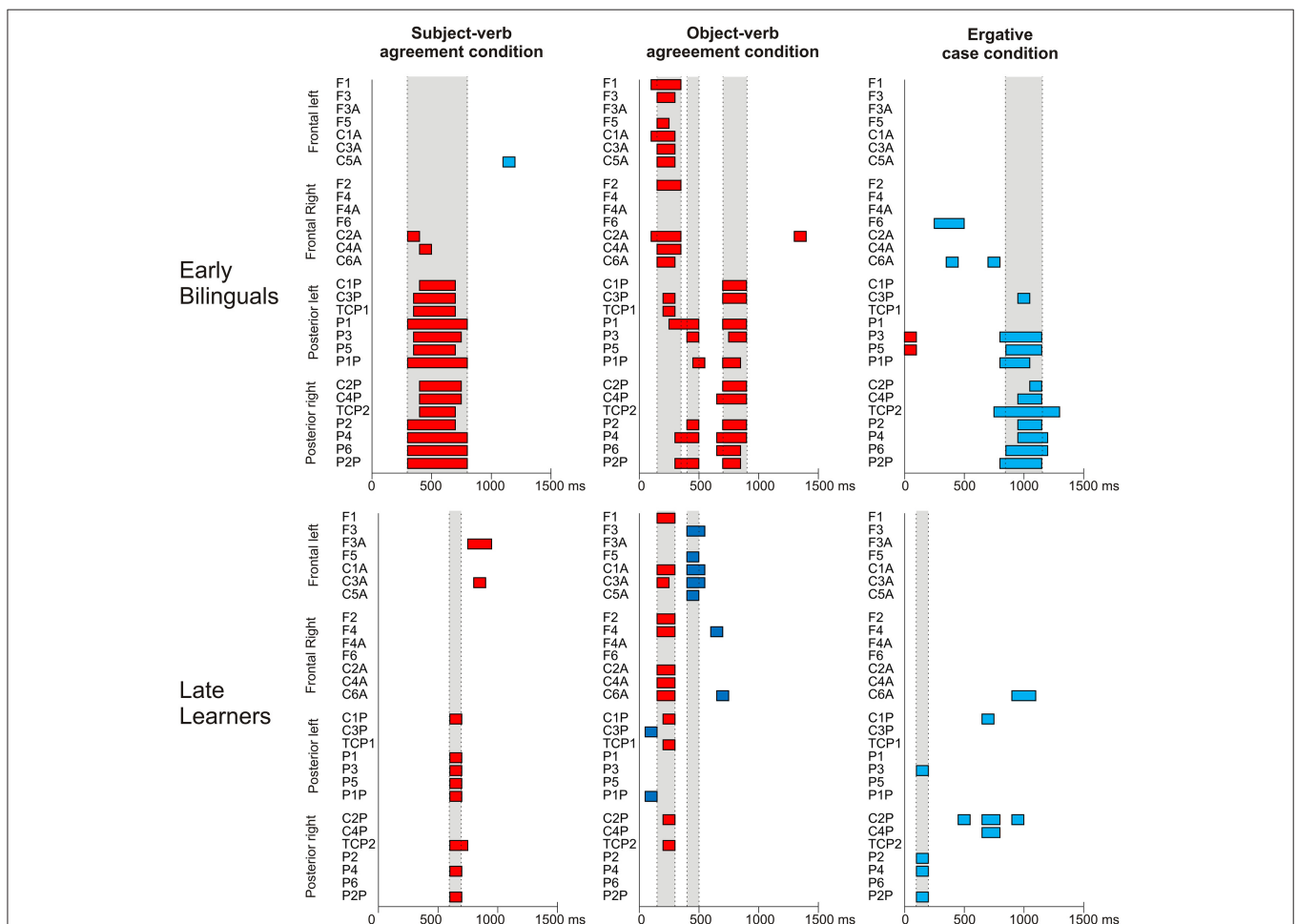
The performance of the selected early bilinguals and late L2 learners was compared for each sentence type (i.e., grammatical, subject-verb agreement violation, ergative case violation, grammatical object and object-verb agreement violation) by means of two-sample *t*-tests on the percentage of hit rates. In addition, the natives' percentage of hit rates was compared to those of early bilinguals and late L2 learners separately by means of two-sample *t*-tests.

### Electrophysiological Data

We used BrainVision Analyzer 2.0 software (Brain Products GmbH, Munich, Germany) to analyze the EEG signal. We used the ocular independent component analyses (Ocular ICA) implemented in BrainVision Analyzer 2.0 Software package (Brain Products) to correct eye movements. We automatically rejected offline those EEG epochs in which any channel either exceeded  $\pm 100 \mu\text{V}$ , had an activity below  $0.5 \mu\text{V}$ , or showed voltage step/sampling above  $50 \mu\text{V}$  within intervals of 200 ms.

Both correctly and incorrectly answered trials were included in the analyses to have similar number of epochs for both groups of participants. For the subject- and object-agreement conditions, the epochs were time-locked to the onset of the auxiliary verb in the grammatical and ungrammatical sentences. For the ergative conditions, the epochs were time-locked to the onset of the ergative marker in the second noun in the grammatical and ungrammatical sentences. All epochs included a pre-stimulus baseline of 100 ms and were 1600 ms long. Overall, 10.78% of the trials were rejected from the analysis for the late L2 learners and 6.45% for the early bilinguals. Subsequent independent-samples *t*-tests on the percent of rejected trials for each sentence type separately showed no significant differences between groups (all *p*-values > 0.05). Baseline was corrected and the linear DC Detrend procedure was performed on the individual segments. ERPs were averaged separately for each participant and sentence type.

First, the ERP pattern for each group and condition (subject-verb agreement, object-verb agreement and ergative case) was analyzed separately. We determined the onsets and durations of the ERP effects by means of *t*-tests on 30 successive time windows of 50 ms that compared grammatical and ungrammatical sentences from 0 to 1500 ms at each electrode using Matlab (R2013b, The MathWorks, Inc., MA, USA). Following previous studies, we controlled for false positives that can occur when a large number of statistical comparisons are performed by considering only those effects that were significant in at least two consecutive 50-ms intervals as reliable (Gunter et al., 1997, 2000; Hahne and Friederici, 2001; Díaz et al., 2011). In addition, the onsets and offsets of the effects were set when at least four electrodes showed significant differences between the grammatical and ungrammatical sentences between the given onsets and offsets (Figure 2).



**FIGURE 2 | Results of the *t*-tests on 50-ms consecutive intervals comparing grammatical and ungrammatical sentences at each electrode and for each condition for early and late bilinguals.** The beginning of the epochs are time-locked to the onset of the critical words (i.e., the auxiliary verb for the subject- and object-agreement conditions and the ergative case marker of the second nominal phrase for the ergative case condition). Significant differences between the grammatical and ungrammatical sentences are indicated by the color bars: Red bars correspond to positive effects and blue bars correspond to negative effects. Discontinuous vertical lines mark the onset and offsets of the significant periods. Gray areas indicate the significant time windows.

For each condition and significant time window, the groups were compared by means of repeated measures ANOVAs on the mean voltages with the within-subjects factors “Grammaticality,” “Region,” “Hemisphere,” and the between-subjects factor “Bilingual Group.” Effects involving the factor “Grammaticality” (main effect and interactions) and the interaction “Grammaticality” × “Bilingual Group” were of interest. Whenever “Grammaticality” and “Bilingual Group” interacted with “Region” and/or “Hemisphere,” separate ANOVAs were performed to test the interaction of the factors “Grammaticality” × “Bilingual Group” for the particular scalp area. Significance levels of the *F*-ratios did not need to be adjusted with the Greenhouse-Geisser correction as all main effects and interactions had only one degree of freedom in the numerator. These analyses were performed using IBM SPSS Statistics 19 (SPSS Inc., Chicago, IL, USA).

## RESULTS

### Behavioral Data

The global hit score was 80.82% ( $\pm 17.29\%$ ) for the early bilinguals and 66.74% ( $\pm 13.97\%$ ) for the late L2 learners. Both groups showed large individual variability in their accuracy (Figure 1), with the scores ranging from high accuracy (98 and 92% hits for early and late groups, respectively) to very poor (49 and 41% hits for early and late groups, respectively). Ten late learners and five early bilinguals showed very poor performance (below 60% hits). Among late learners, the twelve participants that scored globally above 70% were significantly above chance ( $p < 0.05$ ) and one participant with 69% of hits performed marginally ( $p = 0.069$ ) above chance. These thirteen late L2 learners were considered to have sufficient accuracy levels and were included in the ERP analyses. To match the groups for number of participants, the 13 early bilinguals with the highest global accuracy were included in the ERP analysis.

The subgroup of early bilinguals included in the ERP analysis was more accurate in the grammatical judgment task for all

sentences type than the subgroups of late L2 learners (Tables 3, 4). Behaviorally, early bilinguals performed similar to natives (data from Díaz et al., 2011) in all sentence types, although they showed a trend toward poorer performance in the object-verb agreement violations. Late learners performed worse than natives in all sentence types (Tables 3, 4).

### Electrophysiological Data

Figure 2 displays the latencies and durations of the ERP effects revealed by the analysis of the 50-ms intervals for each experimental condition and group of participants. Figure 3 displays the grand average waveforms for the two groups of participants and each violation type against the corresponding grammatical condition. In Table 5, significant effects are reported for the ANOVAs run comparing late and early bilinguals with their corresponding *F*- and *p*-values.

#### Subject-Verb Agreement Condition

The statistical comparisons on the 50-ms consecutive windows revealed a positive effect on overlapping time windows, between 300 and 800 ms for early bilinguals and between 600 and 700 ms for late L2 learners (Figure 2). We compared the groups only for the time window in which both groups coincided, i.e., 600–700 ms, to avoid the analysis of the same data in several ANOVAs.

The ANOVA comparing the two groups in the 600–700 ms window showed a significant effect of “Grammaticality,” an interaction between “Grammaticality” × “Bilingual group” × “Region,” an interaction between “Grammaticality” and “Region” and an interaction between “Grammaticality” × “Bilingual group” × “Region” × “Hemisphere” (Table 5). Because of the 4-way interaction, further ANOVAs were run with the factors “Grammaticality” and “Bilingual group” separately for each area.

The ANOVAs showed only a significant “Grammaticality” effect in the two posterior areas [posterior left:  $F_{(1, 24)} = 20.01$ ,  $p < 0.001$ ; posterior right:  $F_{(1, 24)} = 24.45$ ,  $p < 0.001$ ]. No effects were significant in the frontal areas. Thus, the posterior positivity elicited by the subject-verb agreement violation in the 600–700 ms window was similar between the groups.

#### Object-Verb Agreement Condition

The analysis on the 50-ms time windows showed a positive effect for both groups between 150 and 350 ms, a positivity for early bilinguals and a negativity for late L2 learners between 400 and 500 ms, and a positivity for early bilinguals between 700 and 900 ms (Figure 2).

The ANOVA comparing the two groups in the 150–350 ms window revealed a significant main effect of “Grammaticality” that did not interact with “Bilingual Group,” “Region,” or “Hemisphere” (Table 5). The grammatical violation elicited a similar broadly distributed positivity in both groups.

In the 400–500 ms window, a significant interaction “Grammaticality” × “Bilingual Group” reached significance, as well as the interactions “Grammaticality” × “Region” and “Grammaticality” × “Hemisphere” (Table 5). The interaction “Grammaticality” × “Bilingual group” was analyzed by running separate ANOVAs for each group. For early bilinguals, no effect

**TABLE 3 | Mean percentages of correct responses of natives, early bilinguals, and late L2 learners for each experimental condition.**

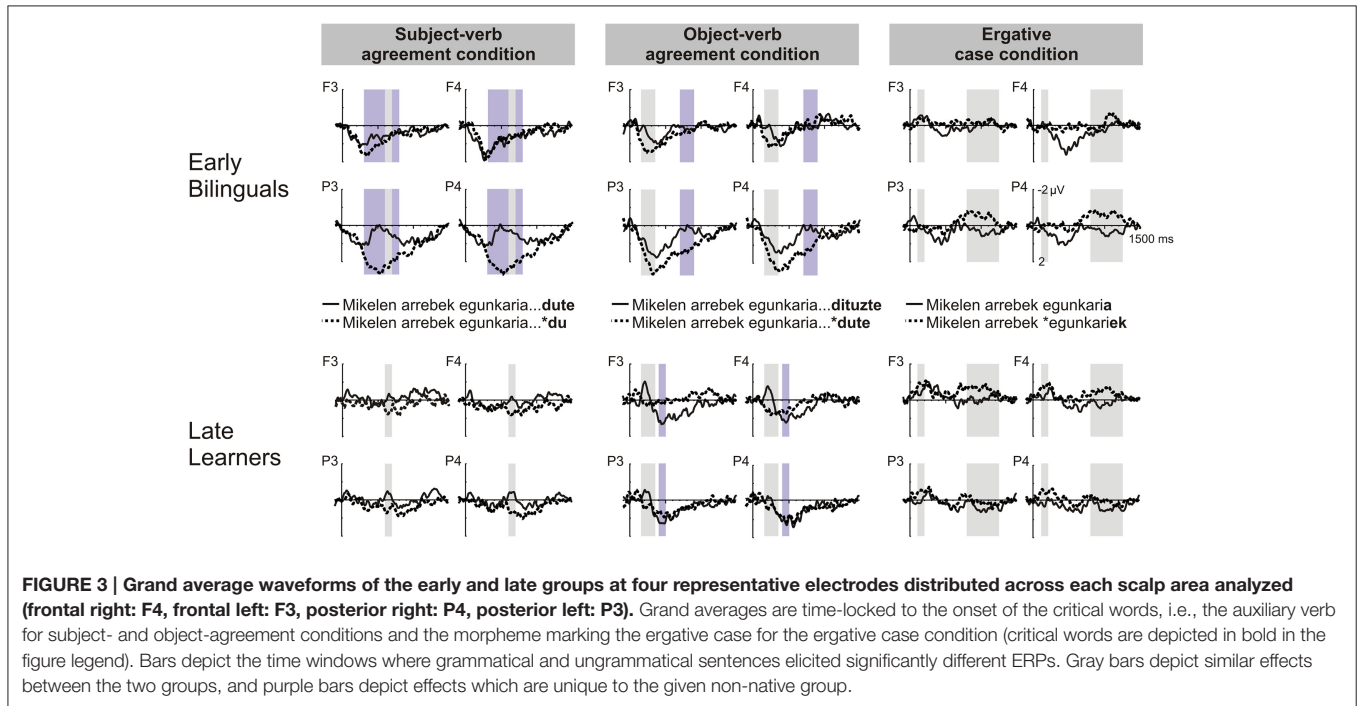
Sentence type	Natives	Early bilinguals	Late L2 learners
Grammatical	95.10 (4.57)	95.38 (3.51)	81.34 (8.39)
Subject-verb agreement violation	91.77 (15.41)	98.07 (2.72)	76.73 (13.78)
Ergative case violation	95.20 (5.04)	95.76 (4.71)	74.23 (7.93)
Grammatical object	93.95 (5.70)	95.00 (5.20)	82.50 (10.30)
Object-verb agreement violation	92.91 (7.46)	88.65 (5.16)	70.00 (15.27)
Total	93.78 (7.63)	94.57 (2.69)	76.96 (6.34)

Standard deviations are in parentheses. The data from a group of native listeners (Díaz et al., 2011) is shown for the sake of comparison.

**TABLE 4 | Two-sample *t*-test comparisons on the mean percentages of correct responses for each experimental condition.**

Sentence type	Early bilinguals vs. natives	Late L2 learners vs. natives	Early bilinguals vs. late L2 learners
Grammatical	$t_{(35)} < 1$	$t_{(35)} = 6.49, p < 0.001$	$t_{(24)} = 5.56, p < 0.001$
Subject- verb agreement violation	$t_{(35)} = 1.45, p > 0.05$	$t_{(35)} = 2.93, p < 0.05$	$t_{(24)} = 8.41, p < 0.001$
Ergative case violation	$t_{(35)} < 1$	$t_{(35)} = 9.84, p < 0.001$	$t_{(24)} = 5.47, p < 0.001$
Grammatical object	$t_{(35)} < 1$	$t_{(35)} = 4.37, p < 0.001$	$t_{(24)} = 3.90, p < 0.001$
Object-verb agreement violation	$t_{(35)} = 1.82, p = 0.076$	$t_{(35)} = 6.16, p < 0.001$	$t_{(24)} = 4.17, p < 0.001$

All groups of participants were compared to each other in pairs. The data from a group of native listeners (Díaz et al., 2011) are analyzed for the sake of comparison.



**FIGURE 3 | Grand average waveforms of the early and late groups at four representative electrodes distributed across each scalp area analyzed (frontal right: F4, frontal left: F3, posterior right: P4, posterior left: P3).** Grand averages are time-locked to the onset of the critical words, i.e., the auxiliary verb for subject- and object-agreement conditions and the morpheme marking the ergative case for the ergative case condition (critical words are depicted in bold in the figure legend). Bars depict the time windows where grammatical and ungrammatical sentences elicited significantly different ERPs. Gray bars depict similar effects between the two groups, and purple bars depict effects which are unique to the given non-native group.

**TABLE 5 | Effects yielded by the ANOVAs on the mean ERP amplitudes comparing the early and late groups for all three conditions separately and for each significant time window revealed by the 50-ms interval analyses. For the sake of completeness, trends toward significant effects are shown but not further analyzed.**

Subject-verb agreement condition [effect: $F_{(1, 12)}$ ]	Object-verb agreement condition [effect: $F_{(1, 12)}$ ]	Ergative case condition [effect: $F_{(1, 12)}$ ]
<u>Time window: 600–700 ms</u> G: 13.01* G × R: 8.61* G × B × R: 6.02* G × B × R × H: 6.55*	<u>Time window: 150–350 ms</u> G: 14.81* <u>Time window: 400–500 ms</u> G × B: 7.80* G × R: 3.34* G × H: 4.36* <u>Time window: 700–900 ms</u> G × B: 4.68* G × R: 5.19*	<u>Time window: 100–200 ms</u> G: 4.12+ G × R: 3.36# <u>Time window: 800–1150 ms</u> G: 8.42* G × B × R: 3.62*

G = Grammaticality, R = Region, H = Hemisphere, B = Bilingual Group.  
\* $p < 0.05$ , + $p = 0.053$ , # $p = 0.079$ .

was significant. For late L2 learners, there was a significant “Grammaticality” effect [ $F_{(1, 12)} = 8.82, p < 0.05$ ] that did not interact with any other factor, hence revealing a broad negativity for ungrammatical sentences.

In the later time window, 700–900 ms, the interactions “Grammaticality” × “Bilingual Group” and “Grammaticality” × “Region” reached significant levels (Table 5). Paired *t*-tests comparing the amplitude for grammatical and ungrammatical

sentences separately for each group showed a trend toward a significant “Grammaticality” effect for early bilinguals [ $t_{(12)} = 1.97, p = 0.071$ ]. This effect was caused by a broad positivity elicited by the grammatical violations. For late L2 learners, there was no significant effect.

### Ergative Case Condition

The analysis on the 50-ms time windows showed a negative effect between 800 and 1150 ms for early bilinguals and between 100 and 200 ms for late L2 learners (Figure 2).

The ANOVAs comparing the two groups in the 100–200 ms window only revealed a trend toward a significant “Grammaticality” effect that did not interact with “Bilingual Group” (Table 5). Hence, the effect of grammaticality was reliable for this time window.

The ANOVA comparing the two groups in the 800–1150 ms window showed a main effect of “Grammaticality” and a trend toward a significant interaction “Grammaticality”  $\times$  “Bilingual group”  $\times$  “Region.” The triple interaction was analyzed despite being only a trend because it involved the group factor. Further ANOVAs for frontal and posterior regions separately with the factors “Grammaticality” and “Bilingual group” revealed only a main effect of “Grammaticality” in both regions [frontal:  $F_{(1, 24)} = 4.28, p < 0.05$ ; posterior:  $F_{(1, 24)} = 11.28, p < 0.01$ ] but no “Grammaticality”  $\times$  “Bilingual group” interaction. The lack of such an interaction showed that both groups displayed similar broad negativities when processing grammatical violations.

## DISCUSSION

The present study compared the brain responses of early bilinguals with high proficiency and late L2 learners with intermediate proficiency (completing a B2 level, Common European Framework) to L2 (Basque) syntactic traits that differed from the participants’ L1 (Spanish). The two groups displayed different ERP responses for the agreement conditions but a similar ERP pattern for the ergative condition. As in native listeners (Díaz et al., 2011), subject-verb agreement violations elicited a posterior positivity between 300 and 800 ms in early bilinguals and a short posterior positivity between 600 and 700 ms in late L2 learners. The latency and polarity of the effect coincide with those of the P600, an index of controlled syntactic and reanalysis repair, which has been consistently reported in native listeners for subject-verb agreement violations across many different languages (Coulson et al., 1998a,b; Hahne and Friederici, 1999; Silva-Pereyra and Carreiras, 2007; Díaz et al., 2011). These findings are in agreement with previous studies showing a native-like P600 in response to similar L1–L2 traits for non-native listeners that attained high levels of proficiency, even when the L2 was acquired late in life. Yet, previous studies reported delayed P600 effects for late L2 learners with intermediate proficiency (Rossi et al., 2006; Kotz et al., 2008; Tanner et al., 2013). Thus, the present L2 subject-verb agreement violation triggered syntactic reanalysis and repair processes in both groups of participants but, in the case of late learners, these syntactic reanalysis and repair processes seem to be slower, given

the delayed latency of the P600, and shallower, given the short duration of the effect.

For the object-verb agreement violations, both groups displayed a similar broad positivity between 150 and 350 ms that was followed by a broad negativity between 400 and 500 ms in late learners and by a marginally significant positivity between 700 and 900 ms in early bilinguals. Using exactly the same procedures, we previously found an early posterior negativity between 200 and 300 ms and a P600 for object-verb agreement violations in natives (Díaz et al., 2011). This early negativity was interpreted as an N200 component reflecting the violation of phonological expectations. Other studies investigating the processing of object-verb agreement violations in natives have reported an N400-P600 pattern (Zawiszewski and Friederici, 2009; Zawiszewski et al., 2011). The N400 component, classically elicited by unexpected words given a semantic context (Kutas and Federmeier, 2000), has also been showed to index conflicts in thematic role assignment in case violations for languages with overt case marking, such as German, Hindi and Basque (Frisch and Schlesewsky, 2001, 2005; Mueller et al., 2005, 2007; Choudhary et al., 2009; Zawiszewski et al., 2011). Given the association of the N400 to thematic processes, the elicitation of an N400 by object-verb agreement violations was claimed to reflect the establishment of thematic roles during verb agreement computations that involve more than one argument.

Given the previous finding of negativities preceding the P600 for object-verb agreement violations, the early positivity in non-native listeners in the present study was unexpected. However, some studies in native German listeners have reported a positivity, rather than a negativity, for thematic computations at verb agreement processing (Mecklinger et al., 1995; Friederici et al., 1998; Bornkessel et al., 2002, 2003). These studies found a centroposterior positivity between 300 and 400 ms, the so-called P345, time-locked to the onset of auxiliary verbs that disambiguated a relative clause toward an object thematic role. Based on these results, it has been claimed that the function indexed by the P345 is a revision of the thematic role assigned to the arguments (Bornkessel et al., 2002). In line with these studies, the early positivity displayed by non-natives in the present study could be triggered by the revision of the initial assignment of the thematic roles of the noun phrases before syntactic repair and reanalysis processes come into play. The fact that no such positivity was elicited in natives by object-verb agreement violations or in non-natives by subject-verb agreement violations could indicate that thematic computations engender a difficulty in non-natives only when agreement involves the object, the agreement relation missing in their L1. Alternatively, the lack of such a positivity in non-native listeners for subject-agreement processing could indicate that different syntactic computations take place depending on the arguments involved (subject or object), as argued for instance in Zawiszewski and Friederici (2009). This latter interpretation of the results would imply that subject- and object-agreement are two different syntactic phenomena. However, the present data alone is not conclusive as to what the underlying process indexed by the early positivity is.

The early positivity for the object condition was followed by a broad negativity in late learners between 400 and 500 ms rather

than a P600. This pattern of results is reminiscent of previous findings with late learners (Osterhout et al., 2006; Guo et al., 2009; McLaughlin et al., 2010; Tanner et al., 2013, 2014). These studies have found an N400, rather than a P600, in novice learners for L2 syntactic violations at the earliest stages of learning but a P600 after 1 year of formal L2 learning in most learners (though there is some individual variation in the timing of the change from an N400 to a P600: Tanner et al., 2014). The N400 in early stages of L2 learning is claimed to index the use of lexical-based heuristics for syntactic processing. The late learners tested in the present study were at early stages of learning. They were in their second year of formal Basque instruction, which makes it likely they are exploiting lexical-semantic aspects rather than syntactic knowledge for accomplishing fast and successful L2 comprehension. In contrast, early bilinguals displayed a marginally significant broad positivity between 700 and 900 ms. We interpret this positivity as an instance of a P600. However, the lack of a posterior distribution and the small amplitude of the positivity suggest that object-verb agreement violations did not trigger native-like reanalysis and repair processes, despite the high proficiency and early AoA of the participants in this group.

This small P600 displayed by early, proficient bilinguals for the object agreement condition contrasts with the native-like P600 effect to object-verb agreement violations reported by Zawiszewski et al. (2011) with a similar population of Spanish early (L1)—Basque (L2) bilinguals who were very proficient in their L2. Three potential sources could be causing, either in isolation or combination, the distinct results reported in Zawiszewski et al. (2011) and the present study. First, the early bilinguals tested in this previous study were similar to natives when detecting object-agreement violations, whereas in the present study, early bilinguals showed a trend toward poorer performance than that of natives in the same task. Second, the use of different sensory modalities for stimuli presentation (written in the previous and auditory in the present study) could be playing a more important role than expected. Despite the fact that sensory modality has been shown not to have an effect on the P600 in native language processing (Hagoort and Brown, 2000; Balconi and Pozzoli, 2005), we cannot rule out that the more demanding auditory presentation, in which word boundaries are not physically present in the stimulus, may be more taxing than visual presentation of isolated words for participants who are not as competent as native listeners. Third, the agreement feature tested in the present and previous study was also different. Zawiszewski et al. (2011) presented person violations, whereas here we presented number violations. It has been shown that subject-person agreement violations engender larger P600 effects than number violations (Nevins et al., 2007; Mancini et al., 2011; Zawiszewski et al., in press). Zawiszewski et al. (in press) studied the Basque native listeners' processing of subject-verb agreement violations of the person feature, number feature, or both. Person violations in all instances elicited a larger P600 than number violations. The same enhancement of the P600 for the person as compared to the number feature has been reported for Hindi and Spanish. Nevins et al. (2007) studied native listeners' processing of subject-verb agreement violations for several features in Hindi: person, number, and gender. A larger P600 was present for feature combinations that encompassed person. Nevins et al.

(2007) concluded that the person feature has a greater salience than other features. Similarly, Mancini et al. (2011) also found a larger and more broadly distributed P600 for person than for number subject-verb agreement violations in Spanish. This difference in the salience of the person and number features agreement may explain why the present group of early bilinguals displays a small P600 effect, while a native-like P600 was reported by Zawiszewski et al. (2011).

The ergative case violation elicited the same brain responses in both groups of non-native listeners: a broad negativity between 800 and 1150 ms. Therefore, neither AoA nor proficiency modulates the ERP responses. The ERP pattern displayed by non-natives was qualitatively different than that of native listeners, who displayed the typical P600 (Díaz et al., 2011). The absence of a P600 for the non-native groups is in line with the findings in Zawiszewski et al. (2011). They found an N400-P600 pattern in natives but only an N400 in early, proficient bilinguals. The N400 preceding the P600 (Frisch and Schlesewsky, 2001, 2005; Mueller et al., 2005, 2007; Zawiszewski et al., 2011) or in isolation (Choudhary et al., 2009) has been repeatedly reported for case violations and is interpreted as indexing processes of thematic assignment. The present negativity does not possess the typical N400 latency (peaks at around 400 ms) or anterior scalp distribution for auditory sentence presentation (Holcomb and Neville, 1990; Connolly and Phillips, 1994; Mueller et al., 2005). However, it is possible that, in non-native listeners, the latency and topography of the ERP components, frequently delayed and broadly distributed as reported for the P600 (Weber-Fox and Neville, 1996; Hahne, 2001; Zawiszewski et al., 2011), does not correspond to those elicited by native listeners. It remains unclear whether the function of the present negativity is analogous to the one reported in natives, which is associated to thematic assignment processes (Frisch and Schlesewsky, 2001, 2005; Mueller et al., 2005, 2007; Choudhary et al., 2009; Zawiszewski et al., 2011), or to the one reported in novice learners instead of P600s, which indexes the use of lexical heuristics (Osterhout et al., 2006; Guo et al., 2009; McLaughlin et al., 2010; Tanner et al., 2013, 2014). We favor the latter interpretation, given that the group of natives tested with the same procedures did not display an N400 (Díaz et al., 2011).

One limitation of the present study is the small sample size, given the very low proficiency in the experimental task of several late learners. The sample size might reduce the sensitivity when trying to capture differences between the groups. Nevertheless, we were able to assess reliable ERP effects within and between the groups. Our results suggest that processing L1–L2 similar traits, like verb-agreement, engages native-like responses (i.e., a P600) in highly proficient, early bilinguals. We interpret this pattern of results as an indication that the presence of verb agreement in the participants' L1 allows verb agreement in L2 to be processed in a native-like fashion, independently of which core arguments are involved in the agreement relation, when the L2 is learned early in life and high proficiency is attained (Zawiszewski et al., 2011). The reduced P600 displayed by proficient, early bilinguals for the object agreement condition together with the non-native early positivity suggests an increased difficulty in applying the L1 processing routines of subject-verb agreement to the processing of the L2 object-agreement. The increased difficulty

in processing object agreement, as compared to subject-verb agreement, is further corroborated by the lack of a P600 effect in late L2 learners. Overall, for the agreement conditions, highly proficient, early bilinguals approximated native processing to a greater extent than late learners with intermediate proficiency. This suggests that the processing of L1–L2 converging traits is influenced by AoA and proficiency. However, the L2-only trait, the ergative case condition, elicited a similar response in both L2 groups: a delayed and broad N400 that was qualitatively different to that of natives, i.e., a P600. This finding indicates that neither AoA nor proficiency influences the brain responses to syntactic traits that are unique to L2. Thus, the comparison of the results between the non-native groups and across the different grammatical traits tested suggests that L1–L2 similarity plays a major role in the neural mechanisms engaged in L2 syntactic processing. The computation of L2 syntactic dependencies engages neural mechanisms that are already present in L1 processing, and the degree to which the pre-existing L1 neural routines can be successfully exploited in the processing of the L2 is influenced by AoA and/or proficiency. In sharp contrast, the processing of syntactic traits that are unique to L2 requires the implementation of new neural routines which do not depend on the L2 age of acquisition, at least when sufficient proficiency is attained in the L2. Nevertheless, the underlying neural processes do not seem to involve native-like processes, even in case of early AoA and high proficiency. Future studies comparing other language pairs are needed to evaluate the cross-linguistic validity of the present findings.

## REFERENCES

- Aleman Bañon, J., Fiorentino, R., and Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: an event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Lang. Res.* 30, 275–306. doi: 10.1177/0267658313515671
- Balconi, M., and Pozzoli, U. (2005). Comprehending semantic and grammatical violations in Italian. N400 and P600 comparison with visual and auditory stimuli. *J. Psycholinguist. Res.* 34, 71–98. doi: 10.1007/s10936-005-3633-6
- Bornkessel, I., Schlesewsky, M., and Friederici, A. D. (2002). Beyond syntax: language-related positivities reflect the revision of hierarchies. *Neuroreport* 13, 361–364. doi: 10.1097/00001756-200203040-00022
- Bornkessel, I., Schlesewsky, M., and Friederici, A. D. (2003). Eliciting thematic reanalysis effects: the role of syntax-independent information during parsing. *Lang. Cogn. Process.* 18, 269–298. doi: 10.1080/01690960244000018
- Caffarra, S., Molinaro, N., Davidson, D., and Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: an empirical review. *Neurosci. Biobehav. Rev.* 51, 31–47. doi: 10.1016/j.neubiorev.2015.01.010
- Chen, L., Shu, H., Liu, Y., Zhao, J., and Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism* 10, 161–174. doi: 10.1017/S136672890700291X
- Choudhary, K. K., Schlesewsky, M., Roehm, D., and Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretively relevant linguistic rules: evidence from Hindi. *Neuropsychologia* 47, 3012–3022. doi: 10.1016/j.neuropsychologia.2009.05.009
- Connolly, J. F., and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266. doi: 10.1162/jocn.1994.6.3.256
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166119
- Corbett, G. G. (1994). “Gender and gender systems,” in *The Encyclopedia of Language and Linguistics*, ed R. Asher (Oxford: Pergamon Press), 1347–1353.
- Coulson, S., King, J. W., and Kutas, M. (1998a). ERPs and domain specificity: beating a straw horse. *Lang. Cogn. Process.* 13, 653–672.
- Coulson, S., King, J. W., and Kutas, M. (1998b). Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58.
- De Rijk, R. P. G. (2007). *Standard Basque, A Progressive Grammar*. Cambridge, MA: MIT Press.
- Díaz, B., Sebastián-Gallés, N., Erdocia, K., Mueller, J. L., and Laka, I. (2011). On the cross-linguistic validity of electrophysiological correlates of morphosyntactic processing: a study of case and agreement violations in Basque. *J. Neurolinguistics* 24, 357–373. doi: 10.1016/j.jneuroling.2010.12.003
- Dowens, M. G., Guo, T., Guo, J., Barber, H., and Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish - evidence from event related potentials. *Neuropsychologia* 49, 1651–1659. doi: 10.1016/j.neuropsychologia.2011.02.034
- Dowens, M. G., Vergara, M., Barber, H. A., and Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *J. Cogn. Neurosci.* 22, 1870–1887. doi: 10.1162/jocn.2009.21304
- Erdocia, K., Zawiszewski, A., and Laka, I. (2014). Word order processing in a second language: from VO to OV. *J. Psycholinguist. Res.* 43, 815–837. doi: 10.1007/s10936-013-9280-4
- 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/S13667289100012X
- Friederici, A. D., Steinhauer, K., Mecklinger, A., and Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biol. Psychol.* 47, 193–221. doi: 10.1016/S0301-0511(97)0033-1
- Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 529–534. doi: 10.1073/pnas.012611199

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

This work was supported by the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007–2013) under REA grant agreement no. 32867 and a postdoctoral fellowship from the Spanish Government (Juan de la Cierva fellowship JCI-2012-12678) to BD, grants from the European Community’s Seventh Framework Programme (FP7/2007-2013): FP7 Cooperation SSH grant agreement no. 613465 (AThEME) awarded to IL and NSG, from the Spanish Ministerio de Economía y Competitividad (PSI2015-66918-P, PSI2012-34071, and PSI2015-71683-REDC) and the Catalan Government (SGR 2014–1210) awarded to NSG, a Ramón y Cajal fellowship (RYC-2010-06520) from the Spanish Government and a grant from the University of the Basque Country (EHUA13/39) to KE, and a Basque Government IT665-13 (2013–2018) and a Spanish Government GRAMMARINPROCESS FFI2012-31360 (2013–2015) awarded to IL. NSG received the prize “ICREA Acadèmia” for excellence in research, funded by the Generalitat de Catalunya. The authors want to thank Xavier Mayoral for his technical support.



- Frisch, S., and Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *Neuroreport* 12, 3391–3394. doi: 10.1097/00001756-200110290-00048
- Frisch, S., and Schlesewsky, M. (2005). The resolution of case conflicts from a neurophysiological perspective. *Cogn. Brain Res.* 25, 484–498. doi: 10.1016/j.cogbrainres.2005.07.010
- Greenberg, J. H. (ed.). (1978). “How does a language acquire gender markers?” in *Universals of Human Language: Word structure* (Stanford, CA: Stanford University Press), 47–82.
- Gunter, T. C., Friederici, A. D., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568. doi: 10.1162/089892900562336
- Gunter, T. C., Stowe, L. A., and Mulder, G. (1997). When syntax meets semantics. *Psychophysiology* 34, 660–676. doi: 10.1111/j.1469-8986.1997.tb02142.x
- Guo, J., Guo, T., Yan, Y., Jiang, N., and Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *J. Neurolinguistics* 22, 123–134. doi: 10.1016/j.jneuroling.2008.09.001
- Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *J. Psycholinguist. Res.* 30, 251–266. doi: 10.1023/A:1010490917575
- Hahne, A., and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *J. Cogn. Neurosci.* 11, 194–205. doi: 10.1162/089892999563328
- Hahne, A., and Friederici, A. D. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism* 4, 123–141. doi: 10.1017/S1366728901000232
- Holcomb, P. J., and Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Process.* 5, 281–312. doi: 10.1080/01690969008407065
- Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain Lang.* 109, 68–74. doi: 10.1016/j.bandl.2008.06.002
- Kotz, S. A., Holcomb, P. J., and Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychol.* 128, 514–527. doi: 10.1016/j.actpsy.2007.10.003
- Kutas, M., and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470. doi: 10.1016/S1364-6613(00)01560-6
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York, NY: Wiley.
- Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject-verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., et al. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Lang. Learn.* 60, 123–150. doi: 10.1111/j.1467-9922.2010.00604.x
- Mecklinger, A., Schriefers, H., Steinhauer, K., and Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: an analysis with event-related potentials. *Mem. Cognit.* 23, 477–494. doi: 10.3758/BF03197249
- Mueller, J. L., Hahne, A., Fujii, Y., and Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *J. Cogn. Neurosci.* 17, 1229–1244. doi: 10.1162/0898929055002463
- Mueller, J. L., Hirotani, M., and Friederici, A. D. (2007). ERP evidence for different strategies in the processing of case markers in native speakers and non-native learners. *BMC Neurosci.* 8:18. doi: 10.1186/1471-2202-8-18
- Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058
- Ojima, S., Nakata, H., and Kakigi, R. (2005). An ERP study of second language learning after childhood: effects of proficiency. *J. Cogn. Neurosci.* 17, 1212–1228. doi: 10.1162/0898929055002436
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Osterhout, L., McLaughlin, J., Pitkänen, 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.1467-9922.2006.0361.x
- Pakulak, E., and Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *J. Cogn. Neurosci.* 23, 2752–2765. doi: 10.1162/jocn.2010.21586
- Pallier, C., Dupoux, E., and Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behav. Res. Methods Instrum. Comput.* 29, 322–327. doi: 10.3758/BF03200583
- Rossi, S., Gugler, M. F., Friederici, A. D., and Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *J. Cogn. Neurosci.* 18, 2030–2048. doi: 10.1162/jocn.2006.18.12.2030
- Sabourin, L., and Stowe, L. A. (2008). Second language processing: when are first and second languages processed similarly? *Second Lang. Res.* 24, 397–430. doi: 10.1177/0267658308090186
- Silva-Pereyra, J. F., and Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Res.* 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029
- Tanner, D., Inoue, K., and Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism* 17, 277–293. doi: 10.1017/S1366728913000370
- Tanner, D., McLaughlin, J., Herschensohn, J., and Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism* 16, 367–382. doi: 10.1017/S1366728912000302
- Tokowicz, N., and MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: an event-related potential investigation. *Stud. Second Lang. Acquisit.* 27, 173–204. doi: 10.1017/S0272263105050102
- UNESCO World Report: Investing in Cultural Diversity and Intercultural Dialogue. (2009). Luxembourg: UNESCO.
- Van Hell, J. G., and Tokowicz, N. (2010). Event-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 proficiency levels. *Second Lang. Res.* 26, 43–74. doi: 10.1177/0267658309337637
- V Encuesta Sociolingüística: 2011 (2003). Vitoria-Gasteiz: Gobierno Vasco.
- Weber-Fox, C. M., and Neville, H. J. (1996). Maturational constraints on functional specialization for language processing: ERP and behavioral evidence in bilingual speakers. *J. Cogn. Neurosci.* 8, 231–256. doi: 10.1162/jocn.1996.8.3.231
- World Medical Association. (2013). World medical association declaration of helsinki ethical principles for medical research involving human subjects. *J. Am. Med. Assoc.* 310, 2191–2194. doi: 10.1001/jama.2013.281053
- Zawiszewski, A., and Friederici, A. D. (2009). Processing canonical and non-canonical sentences in Basque: the case of object-verb agreement as revealed by event-related brain potentials. *Brain Res.* 1284, 161–179. doi: 10.1016/j.brainres.2009.05.099
- Zawiszewski, A., Gutierrez, E., Fernandez, B., and Laka, I. (2011). Language distance and non-native syntactic processing: evidence from event-related potentials. *Bilingualism* 14, 400–411. doi: 10.1017/S1366728910000350
- Zawiszewski, A., Santesteban, M., and Laka, I. (in press). Phi-features reloaded: an event-related potential study on person and number agreement processing. *Appl. Psycholinguist.* doi: 10.1017/S014271641500017X

**Conflict of Interest Statement:** 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.

The reviewer RT and the handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Díaz, Erdocia, de Menezes, Mueller, Sebastián-Gallés and Laka. 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) or licensor 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.



# Verbal Semantics Drives Early Anticipatory Eye Movements during the Comprehension of Verb-Initial Sentences

Sebastian Sauppe<sup>1,2,3\*</sup>

<sup>1</sup> Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, <sup>2</sup> Department of Linguistics, Ruhr University Bochum, Bochum, Germany, <sup>3</sup> Department of Linguistics and Information Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Martin John Pickering,  
University of Edinburgh, UK  
Matthew Wagers,  
University of California, Santa Cruz,  
USA

### \*Correspondence:

Sebastian Sauppe  
sebastian.sauppe@mpi.nl

### Specialty section:

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

**Received:** 31 August 2015

**Accepted:** 18 January 2016

**Published:** 09 February 2016

### Citation:

Sauppe S (2016) Verbal Semantics Drives Early Anticipatory Eye Movements during the Comprehension of Verb-Initial Sentences. *Front. Psychol.* 7:95. doi: 10.3389/fpsyg.2016.00095

Studies on anticipatory processes during sentence comprehension often focus on the prediction of postverbal direct objects. In subject-initial languages (the target of most studies so far), however, the position in the sentence, the syntactic function, and the semantic role of arguments are often conflated. For example, in the sentence “The frog will eat the fly” the syntactic object (“fly”) is at the same time also the last word and the patient argument of the verb. It is therefore not apparent which kind of information listeners orient to for predictive processing during sentence comprehension. A visual world eye tracking study on the verb-initial language Tagalog (Austronesian) tested what kind of information listeners use to anticipate upcoming postverbal linguistic input. The grammatical structure of Tagalog allows to test whether listeners’ anticipatory gaze behavior is guided by predictions of the linear order of words, by syntactic functions (e.g., subject/object), or by semantic roles (agent/patient). Participants heard sentences of the type “Eat frog fly” or “Eat fly frog” (both meaning “The frog will eat the fly”) while looking at displays containing an agent referent (“frog”), a patient referent (“fly”) and a distractor. The verb carried morphological marking that allowed the order and syntactic function of agent and patient to be inferred. After having heard the verb, listeners fixated on the agent irrespective of its syntactic function or position in the sentence. While hearing the first-mentioned argument, listeners fixated on the corresponding referent in the display accordingly and then initiated saccades to the last-mentioned referent before it was encountered. The results indicate that listeners used verbal semantics to identify referents and their semantic roles early; information about word order or syntactic functions did not influence anticipatory gaze behavior directly after the verb was heard. In this verb-initial language, event semantics takes early precedence during the comprehension of sentences, while arguments are anticipated temporally more local to when they are encountered. The current experiment thus helps to better understand anticipation during language processing by employing linguistic structures not available in previously studied subject-initial languages.

**Keywords:** sentence comprehension, anticipation, prediction, visual world eye tracking, Tagalog, verb-initial word order

## 1. INTRODUCTION

Anticipation, the prediction of upcoming events, is an important property of human cognition and it has been argued recently that brains are essentially “prediction machines” (Clark, 2013, cf. also Bubic et al., 2010). Predictive processes are found, for example, in interaction between individuals when people predict the outcome of actions performed by others (Sebanz and Knoblich, 2009) and even their movements (Kilner et al., 2004).

Anticipation is also involved in language processing. During the comprehension of spoken or written sentences, language users build predictions about the upcoming linguistic input. Words are, for example, read faster when they are predictable from the context as compared to unpredictable words (Ehrlich and Rayner, 1981). Language users may even predict the phonological form of an upcoming word: DeLong et al. (2005) found differential EEG responses when listeners encountered a determiner (*a/an*) that did not fit with the noun that they assumed will follow (“*The day was breezy so the boy went outside to fly... a kite vs. an airplane*”). Anticipatory processes are also found in conversation where listeners predict the end of their interlocutor’s turn, in order to be able to take their own turn in a timely manner (Magyari and de Ruiter, 2012; Magyari et al., 2014).

The visual world paradigm has been used extensively to investigate predictive processes during language comprehension. In this experimental paradigm, participants see a display and hear an accompanying sentence while their eye movements are recorded (cf. Huettig et al., 2011a for a review). In a seminal visual world study, Altmann and Kamide (1999) showed that in English the lexical semantics of verbs is used to anticipate the syntactic object of a sentence by incrementally narrowing down the set of potential referents. Participants saw displays showing, e.g., a boy, a ball, a toy train, a toy car, and a cake, and heard sentences of the form “*The boy will move/eat the...* ”. The verb of the sentence could either take any of the depicted things (*move*) or only one of them (*eat*) as its syntactic object. Listeners used the verb’s selectional restrictions and fixated on the corresponding element in the display already before it was mentioned when the verb only allowed one object referent in this position (*eat* and *cake* in this case).

Further visual world studies substantiate the idea that sentence comprehension is highly predictive and that listeners use various kinds of information to form their expectations. Kamide et al. (2003b) showed that case marking information can be combined with semantic information from the verb in German to anticipate syntactic objects. Kamide et al. (2003a) showed that information from several constituents can be combined to predict upcoming elements in English ditransitive sentences and in verb-final Japanese sentences. Boland (2005) showed that arguments are more likely to be anticipated than adjuncts in English. Knoeferle et al. (2005) showed that listeners rapidly integrate visual information that is provided to them and that this information is used to anticipate object referents in German, even when the sentences accompanying a display describe unusual situations and therefore run counter to listeners’ world knowledge.

All of these studies have in common that they investigated how information provided by sentential and visual context are integrated to predict elements that occur at the end of sentences. The already encountered input restricts language users’ attention to the anticipation of the only remaining element of the sentence. Transitive verbs, such as *eat*, take two arguments and in languages with subject-initial word order (e.g., English and German), listeners already have heard one of the arguments when they encounter the verb—the point from which anticipatory eye movements are measured in most studies. Thus, listeners already have information about one argument, including its referential identity and its semantic role (in the case of Kamide et al., 2003a even about two arguments of ditransitives). Put differently, in previous studies on subject-initial languages the anticipation target has always been a single element at the end of a sentence, conflating syntactic function, word order, and semantic role.

There is thus still an open question regarding what kind of information listeners orient to for predictive processes during sentence comprehension. Do they try to anticipate referents based on syntactic function (e.g., direct object)? Alternatively, are their expectations based solely on what they expect to follow next? Or do listeners rather exploit semantic information to form expectations about the event and therefore anticipate referents carrying certain semantic roles (e.g., patient or goal)? Unfortunately, studies of subject-initial languages are not suited to answer these questions because the three different types of anticipation targets are conflated on the last noun phrase position that is usually employed to test prediction processes. Taking Altmann and Kamide’s (1999) sentences, *cake* is the direct object, the patient and the word directly following the verb. Examining the prediction of this element cannot differentiate between these three types of information as the anticipation target.

Verb-initial languages offer a possibility to disentangle these various theoretical possibilities. In these languages, the verb is the first word of a sentence and information about the described event and selectional restrictions are provided upfront, potentially enabling listeners to identify referents and the semantic roles that they play. Importantly, the early position of the verb may enable listeners to anticipate upcoming arguments before any of them is mentioned. This means that all three anticipation target types are still available—prediction based on semantic roles, on syntactic functions, or on word order. In subject-initial languages, on the other hand, one argument is always mentioned before the verb.

In the following, a visual world eye tracking experiment on Tagalog will be reported. Tagalog is an Austronesian language primarily spoken in the Philippines. The experiment was devised to test what kind of information listeners anticipate in verb-initial languages upon having heard the verb.

### 1.1. Current Experiment

In the experiment described below, participants looked at visual displays depicting three potential referents (cf. **Figure 1**) while hearing verb-initial Tagalog sentences. Two elements in the display corresponded to the agent and to the patient of the sentences, the third element was an unrelated distractor. Participants’ eye movements were recorded in order to analyze

their looks to the elements as the sentences unfolded. The experiment was designed to investigate what kind of information listeners orient toward upon hearing a sentence-initial verb and what it is that they anticipate, especially when there are more possible anticipation targets than just the last word of the sentence. There are different sentence types in Tagalog that can be used to test the three possible anticipation targets; these sentence types are described in the following.

Basic word order in Tagalog is verb-initial and the verb carries voice affixes that cross-reference the semantic role of one of its arguments. This argument is marked by *ang* and will be referred to as the pivot argument. Non-pivot arguments that do not have their semantic role cross-referenced are marked by *ng*. Canonically and most frequently, the non-pivot argument immediately follows the verb and the pivot argument is realized sentence-finally (cf., e.g., Himmelmann, 2005, p. 357).

In (1a)<sup>1</sup> the agent in the event (*frog*) is marked by *ang* and the verb exhibits voice morphology that signals that the semantic role of this pivot argument is agent (AV). In (1b) the patient (*fly*) is marked by *ang* and the verb signals that the pivot argument's semantic role is patient by means of different voice morphology (PV)<sup>2</sup>.

- (1) a. Kakain sa umaga ng=langaw ang=palaka  
eat:AV in the morning NPVT=fly (P) PVT=frog (A)  
“The frog will eat a fly in the morning.”
- b. Kakainin sa umaga ng=palaka  
eat:PV in the morning NPVT=frog (A)  
ang=langaw  
PVT=fly (P)  
“A/the frog will eat the fly in the morning.”<sup>3</sup>

Importantly, both sentences are equally transitive. Kroeger (1993, pp. 40–48) shows with a number of syntactic tests that *ng*-marked patients in agent pivot sentences (1a) and *ng*-marked agents in patient pivot sentences (1b) are arguments of the verb. Tagalog can thus be described as exhibiting a so-called symmetrical voice system (Foley, 2008; Riesberg, 2014). This is in contrast to English where passive sentences are intransitive and the agent may only be realized as oblique.

Therefore, in sentences like (1), the initial verb provides language users with semantic information about the described event. In the context of a visual world eye tracking experiment, this might allow them to identify which referents in the visual display could sensibly be involved in the described event (e.g., a frog as the agent and a fly as the patient in sentences like in 1 or a boy and a cake as in Altmann and Kamide, 1999).

<sup>1</sup>The following abbreviations are used in the current paper: A, agent; AV, agent voice; NPVT, non-pivot argument; P, patient; PV, patient voice; PVT, pivot argument; RP, recent perfective aspect.

<sup>2</sup>Tagalog also exhibits a variety of other voice forms where, e.g., the instrument, the beneficient or the location of an event is the pivot and has its semantic role cross-referenced at the verb (e.g., Schachter and Otanes, 1972; Himmelmann, 2005).

<sup>3</sup>Differences in the definiteness of agent and patient in the translations arise due to constraints on interpreting the *ang*-marked argument as specific (Adams and Manaster-Ramer, 1988, cf. also Latrouite, 2015).

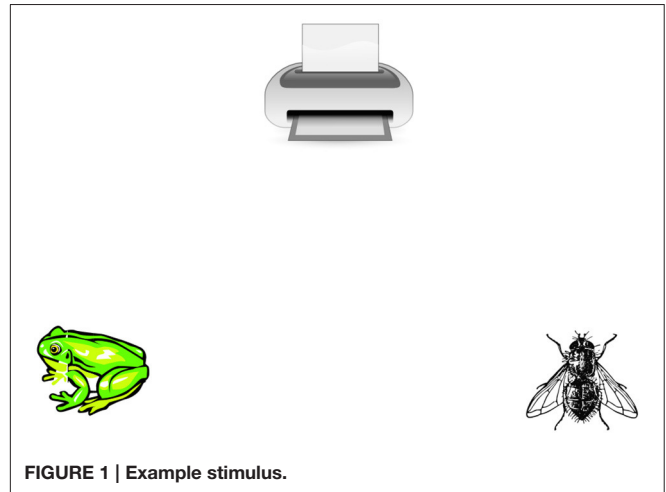


FIGURE 1 | Example stimulus.

Additionally, the voice marking carried by the verb provides information about the canonical order of agent and patient in the unfolding sentence. When the verb signals that the agent is the pivot (example 1a), listeners know that it will be canonically and most frequently realized sentence-finally, i.e., that the canonical order is [patient agent]. When the verb marks a patient pivot (example 1b), listeners know that the canonical order is [agent patient]. Thus, the sentence-initial verb provides listeners with information about the event from which agent and patient referents in the display can be inferred and it provides them with information about the canonical and most frequent order in which these referents will be mentioned.

Tagalog also exhibits a construction that differs from the sentences in (1) in an interesting way. Sentences in the recent perfective aspect describe events that recently happened. In these sentences the verb is not marked for voice but carries an invariant aspect marker. Thus, there is no pivot in recent perfective sentences (2) and the canonical order of arguments is [agent patient].

- (2) Kakakain pa lang ng=palaka sa=langaw  
eat:RP just NPVT=frog (A) NPVT=fly (P)  
“A/the frog just ate the fly.”

Taken together, sentences with agent pivots, patient pivots and recent perfective sentences provide a way of investigating what kind of information language users anticipate after having heard a sentence-initial verb. The three sentence types contrast in their verb marking, i.e., whether the semantic role of a pivot argument is reflected on the verb (1) or not (2)—and if there is a pivot argument, whether it is the agent or the patient of the sentence. Additionally, the three sentence types also differ in the canonical order of the agent and patient arguments ([patient agent] for agent pivot sentences, 1a, and [agent patient] for patient pivot and recent perfective sentences, 1b and 2). Whether Tagalog listeners anticipate upcoming linguistic input on the basis of semantic or syntactic information can be investigated by comparing the comprehension of these three sentence types. It is possible to formulate differential hypotheses for each possible kind of information that may be used in anticipatory processing

based on listeners' behavior during sentence comprehension. These hypotheses will be laid out in more detail in the following.

If Tagalog listeners primarily orient toward syntactic information in anticipation, they could use the semantic and morphosyntactic information provided by the verb to identify agent and patient referents and assign syntactic functions (pivot, non-pivot) to them.

A strong form of syntactically based anticipation would be the prediction of pivot arguments, i.e., that listeners anticipate the sentence-final pivot NP by already fixating on the corresponding referent in the display while or shortly after hearing the sentence-initial verb. When the verb signals that the agent is the pivot (1a), listeners should look toward the agent more after having heard the verb than when the patient is signaled to be the pivot (1b)—in which case listeners should direct their gaze toward the patient. Sauppe et al. (2013) found that in Tagalog sentence production the pivot argument plays an important role early in the planning of sentences: Tagalog speakers select a pivot at the outset of formulation in order to be able to retrieve an appropriate voice affix. If the role of the pivot argument is mirrored in anticipatory processing during sentence comprehension, fixation preferences for the agent in (1a) or the patient in (1b) are expected shortly after listeners encountered the verb.

Another syntactically based process would be the anticipation of the first-mentioned argument upon hearing the verb. Under this scenario, listeners use verbal information to identify referents and their canonical order to determine whether agent or patient will be mentioned first and will subsequently direct their gaze toward them. After having heard a verb that signals an agent pivot, listeners should direct their gaze toward the patient element in the display because the canonical word order for these sentences is [patient agent]. After having heard a verb with patient pivot or recent perfective marking, listeners should direct their gaze toward the agent referent (cf. Table 1).

Finally, if Tagalog listeners directed their attention toward semantic roles and therefore toward the structure of the event, they should fixate on the agent in all three sentence types after having heard the verb. Agents play a prominent role in communication in general because they are initiators of events. Cohn and Paczynski (2013) propose that agents are centrally involved in building representations of events and may take early precedence during the cognition of events since they are the “heads of causal chains that affect patients” (Kemmerer, 2012). Agents are also attended to more than patients by infants (Robertson and Suci, 1980) and play a highlighted role in many grammatical hierarchies (Aissen, 1999; Lockwood and Macaulay, 2012). Given these points, it seems justified to assume that agents

are the target of anticipatory processes in Tagalog if prediction was guided by semantic roles.

In the grammatical literature it has also been proposed that Tagalog exhibits a “patient primacy,” partly because sentences in which the patient is the pivot are more frequent than agent pivot sentences (cf. Latrouite, 2011 for a discussion). Theoretically, the patient could thus also be fixated preferentially after the verb was heard. However, on the hypothesis that the anticipation of semantic roles would mainly serve to construct an event representation, it seems a priori more likely that agents would be targeted for this purpose.

## 2. EXPERIMENT

### 2.1. Participants

Forty-nine students of the University of the Philippines, Diliman, participated in the experiment for payment (mean age = 18.8 years, 22 male). All of them reported being native speakers of Tagalog and speaking the language with at least one of their parents. All participants had normal or corrected-to-normal vision.

The reported experiment conforms to the American Psychological Association's ethical principle of psychologists and code of conduct (as declared by the ombudsman of the Max Planck Institute for Psycholinguistics). Written informed consent was obtained from participants at the beginning of the experiment session.

### 2.2. Materials and Methods

#### 2.2.1. Materials

In the experiment, participants looked at stimulus displays while hearing pre-recorded sentences. Stimulus displays consisted of three colored line drawings that were arranged in a triangular shape (Figure 1). Line drawings either represented the agent or patient of the event described in the accompanying sentence or were distractors which were not mentioned. The position of agent, patient and distractor was counterbalanced across displays.

Displays were paired with sentences that were either intransitive or transitive. All intransitive sentences were fillers. Transitive sentences described a range of animacy scenarios in which agent and patient were humans, animals, or inanimate entities. However, scenarios in which both agents and patients were inanimate were not included. Verbs and arguments were semantically associated to varying degrees (ranging from *police car chases thief to owl carries bag*).

In all sentences the initial verb was followed by an adverb (*sa umaga* “in the morning,” *sa tanghali* “at noon,” or *sa hapon* “in the afternoon” for sentences as in 1 and *pa lang* “just” for recent perfective sentences as in 2). The adverb was included to increase the time between hearing the verb and the first noun phrase<sup>4</sup> in order to give participants time to parse the verb and direct their

TABLE 1 | Overview of sentence types; pivot arguments underlined.

Region	Verb	Adverb	NP1	NP2
Agent pivot	(1a) eat <sub>AV</sub>	in the morning	fly (P)	frog (A)
Patient pivot	(1b) eat <sub>PV</sub>	in the morning	frog (A)	fly (P)
Recent perfective	(2) eat <sub>AP</sub>	just	frog (A)	fly (P)

<sup>4</sup>Strictly speaking, the arguments are expressed by determiner phrases headed by the markers *ng*, *ang* and *sa*, which define the referential meaning of the phrases. Content words are not sub-classified for syntactic categories in Tagalog and therefore there are no noun and verb classes (Himmelmann, 2008). For the sake of simplicity, however, the term NP will be used in this paper, following Himmelmann (2005).

gaze toward the anticipation target (cf., e.g., Kamide et al., 2003b; Mishra et al., 2012 for similar stimulus sentence structures).

Sentences were recorded by a female native speaker of Tagalog and had a neutral intonation contour so that none of the words was particularly highlighted.

Fifty-one critical displays were paired with transitive sentences which exhibited either marking of agent voice, patient voice, or recent perfective on the sentence-initial verb; agent and patient were depicted together with a distractor element semantically unrelated to the two arguments and the verb. In these displays only one element could be the agent referent and only one could be the patient referent. Seventy-nine filler displays depicted only one argument of the accompanying sentence and two distractors. The sentences were either intransitive and thus included only one argument (49 sentence-display pairs) or transitive (30 sentence-display pairs). In the latter case, one argument was mentioned but not depicted as an element in the display or two elements were possible agents or patients of the verb. Three pseudo-randomized lists were created so that each critical display occurred with one of the three sentence types in each list and at least one filler intervened between any two critical displays. For sentences describing scenarios where humans were acted on, either undergoer voice or recent perfective was used in two lists as there is a grammatical constraint against agent voice when the patient is human (Latrouite, 2011).

## 2.2.2. Procedure

Participants were seated in front of a 17" laptop computer with a screen resolution of 1024 × 768 pixels. Eye movements were recorded with 120 Hz sampling rate by a SMI RED-m eye tracker attached below the computer's screen. Auditory stimuli were presented via headphones.

Trials began with the presentation of a fixation cross in the middle of the screen that triggered the presentation of the experimental display after participants looked at the cross for 700 ms. The auditory presentation of sentences started 1000 ms after the onset of the display, which stayed visible until after the end of the sentence.

After a quarter of the trials participants were asked to indicate whether all the referents mentioned in the sentence were also depicted; this was always true for the critical transitive sentences and sometimes true and sometimes false for filler sentences. Five practice trials were included at the beginning of the experiment.

The judgment task that participants had to carry out was similar to the task employed in Altmann and Kamide (1999) where participants had to indicate whether the event could apply to the picture, which was the case when all relevant referents were depicted. This kind of "look and listen" task was also employed in other visual world eye tracking studies investigating anticipatory processes (e.g., Huettig et al., 2011b). Huettig et al. (2011a, p. 154) conclude that "the listeners' eye movements during a trial of a visual world experiment reflect the direction of their visual attention, which depends both on the visual and auditory input," i.e., listeners look at the elements in the display as they are mentioned and become relevant (Huettig et al., 2011a, p. 153). The linking hypothesis employed in the current paper is thus that

listeners' gaze is a reliable reflection of their attention allocation during sentence comprehension.

Before testing, participants read instructions for the experiment in Tagalog and completed a questionnaire on their linguistic background. The whole session lasted approximately 35 min.

## 2.2.3. Analyses

To test the hypotheses regarding possible anticipation targets outlined above, the time course of participants' fixations to agent and patient referents in experimental displays during the comprehension of the three different sentence types was analyzed.

Likelihoods of agent and patient fixations were analyzed with quasi-logistic linear mixed effects regression models (Pinheiro and Bates, 2000; Barr, 2008; Bates et al., 2015; R Core Team, 2015) in three time windows. The first time window encompassed the sentence-initial verb and the immediately following adverb (*Verb + Adverb* region, duration: mean = 1183 ms, *SD* = 96 ms), the second time window spanned the period during which the first argument was presented (*NP1* region, duration: mean = 703 ms, *SD* = 187 ms), finally the third time window covered the presentation of the second argument (*NP2* region, duration: mean = 815 ms, *SD* = 201 ms). To account for variations in the duration of regions across stimuli due to differing word lengths, the duration of each time window was normalized. For every stimulus, the onset of the respective region for each analysis time window corresponded to *time* = 0 and the region's offset corresponded to *time* = 1. In this way, only fixations that occurred during the presentation of any given sentence region of each item were included in the corresponding analysis time windows. Fixations were aggregated into empirical logits over five consecutive bins for each analysis time window.

Time and sentence type were included as predictors in all regression models and the maximal random effects structure justified by design (that allowed the models to converge) was used (Barr, 2013; Barr et al., 2013). Significance of fixed effects was assessed using Type II Wald *F*-tests with Kenward-Roger approximation of denominator degrees of freedom (Kenward and Roger, 1997; Fox and Weisberg, 2011; Halekoh and Højsgaard, 2014). Sentence type as categorical predictor was coded with Helmert contrasts.

Trials were excluded from analyses if track-loss occurred, defined as the eye tracker having lost the participant's eyes for more than 650 ms (236 trials, 9.4%), or due to technical problems with the recording equipment (15 trials, 0.6%). Trials were also excluded if the question after a given trial was answered incorrectly; six participants that answered less than 80% of questions correctly were excluded entirely from the analyses (296 trials, 11.8%). One item was excluded from analyses because it was accidentally in the same condition in all lists. In one list, the trials from one critical display were excluded because it accidentally was presented together with a filler sentence. Three combinations of display and recent perfective sentence were discarded because they were rated as only marginally acceptable in a *post-hoc* internet-based acceptability rating study conducted with 50 Tagalog speakers from the Philippines (51

trials, 2%). Nine stimuli were excluded because the accuracy of agent recognition (given the display and the voiceless and aspect-less gerund form of the verb) was less than 10% above chance in a *post-hoc* internet-based rating study with 29 Tagalog speakers from the Philippines (322 trials, 13%). In total, 1568 trials were included in the analyses.

### 2.3. Results

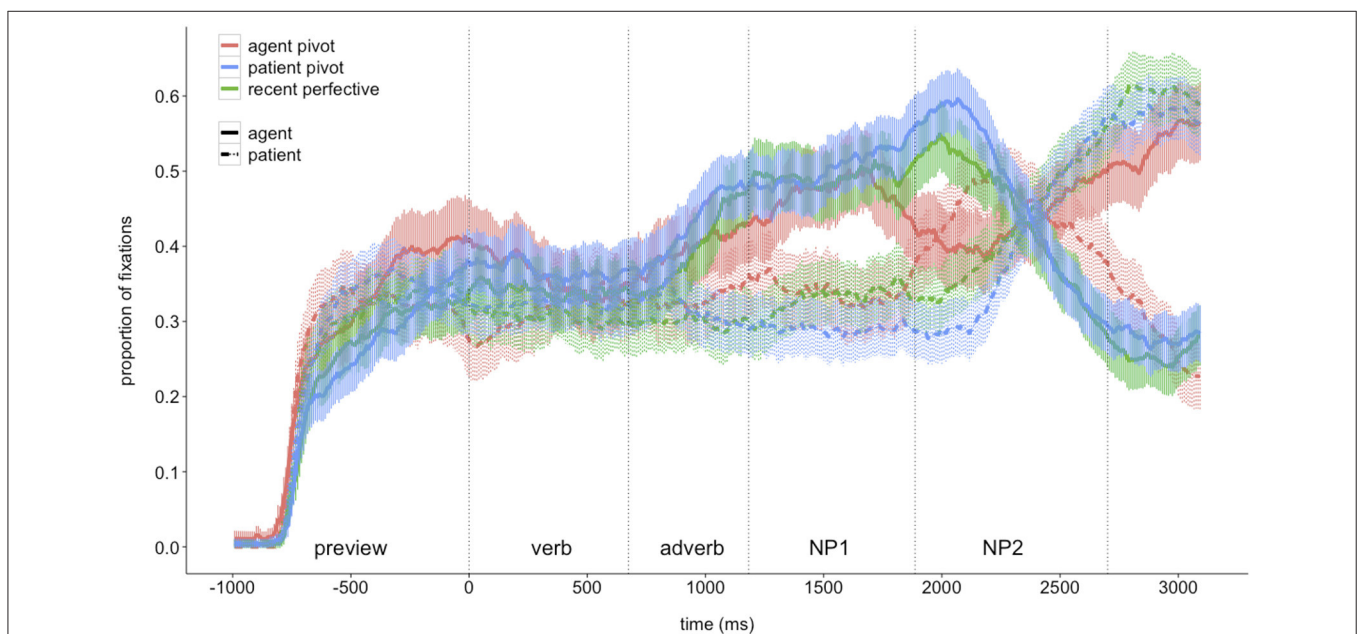
The time course of listeners' fixations to agents and patients during the auditory presentation of the three different sentence types is shown in **Figure 2**. Visual inspection of the graph suggests that agent fixations increased during the *Verb + Adverb* region in all three sentence types after listeners encountered the verb. Agent fixations then continued to increase in sentences with patient voice (1b) and recent perfective marking (2) until the agent was mentioned. For sentences with agent voice marking (1a), participants' agent fixations decreased during the *NP1* region where the patient was mentioned and increased again later when the agent was mentioned during the *NP2* region. In contrast, fixations to the patient did not increase during the *Verb + Adverb* region in any of the sentence types. In sentences where the patient was encountered after the adverb (1a), participants' fixations to that referent started increasing toward the end of the *NP1* region and decreased during the *NP2* region in which the agent was mentioned. In sentences with patient pivots or recent perfective marking the patient was mentioned only sentence-finally. In these sentences, participants' fixations to the patient started to increase only toward the end of the *NP1* region and during the *NP2* region where it was mentioned.

**Table 2** shows the results of the quasi-logistic linear mixed effects regression models for fixations to the agent in the three analysis time windows. During the *Verb + Adverb* region, only

time is a significant predictor. This means that during this time window, the likelihood of agent fixations increased over time and it did so to a similar degree in all sentence types; in other words, the slope does not vary with verb marking.

During the *NP1* region, there was a steeper increase in agent fixations by-subjects in sentences where it was mentioned first, i.e., sentences with a sentence-final patient pivot (1b) or recent perfective marking (2). The fixation patterns associated with these two sentence types were highly similar but differed from fixation patterns observed when listeners heard sentences with an agent pivot (i.e., where the agent was heard first, 1a). This difference arose because agent fixations decreased toward the end of this time window in sentences with sentence-final agent pivots but not in the other two sentence types. By-items, the interaction of time and sentence type did not reach statistical significance. There is, however, a significant main effect of sentence type meaning that there were more fixations to the agent for sentences in which it was mentioned first, i.e., sentences with patient pivot (1b) or recent perfective marking (2), as compared to agent pivot sentences.

During the *NP2* region, agent fixations decreased in sentences with patient pivots and recent perfective marking, in which the patient was mentioned in sentence-final position, as compared to agent pivot sentences with the agent in final position. In fact, fixations to the agent in the latter sentence type increased during this time window. Additionally, there was a steeper decrease in agent fixations for sentences where the patient was the pivot argument (1b) as compared to pivot-less recent perfective sentences in the by-subjects regression model. However, this effect was not detectable in the by-items model.



**FIGURE 2 |** Fixations to depictions of agents and patients during auditory presentation of three different sentence types; ribbons indicate 95% confidence intervals; vertical dotted lines indicate mean onset and offset of regions in the critical sentences (cf. Table 1).

**TABLE 2 |** Quasi-logistic linear mixed effects regression results predicting empirical logits of fixations to the agent referent in three different sentence types.

	By subject				By item			
	$\beta$	SE	95% Wald CI	F statistic	$\beta$	SE	95% Wald CI	F statistic
<i>Verb + Adverb</i>								
Intercept	-0.50	0.05	-0.59, -0.40		-0.48	0.08	-0.63, -0.32	
Time	0.36	0.11	0.14, 0.57	$F_{(1, 10842)} = 14.91^{***}$	0.38	0.16	0.07, 0.70	$F_{(1, 4241)} = 6.28^*$
{PV, RP} vs. AV	<0.01	0.09	-0.17, 0.17	$F_{(2, 5754)} = 1.17$	0.03	0.07	-0.12, 0.17	$F_{(2, 18430)} = 2.23$
PV vs. RP	0.13	0.09	-0.05, 0.32		0.12	0.07	>-0.01, 0.25	
Time $\times$ {PV, RP} vs. AV	0.29	0.29	-0.28, 0.85	$F_{(2, 10546)} = 0.67$	0.17	0.18	-0.19, 0.53	$F_{(2, 43723)} = 0.51$
Time $\times$ PV vs. RP	0.15	0.19	-0.21, 0.52		0.09	0.25	-0.40, 0.57	
<i>NP1</i>								
Intercept	-0.05	0.05	-0.15, -0.05		-0.06	0.07	-0.19, 0.08	
Time	0.15	0.10	-0.04, 0.34	$F_{(1, 3879)} = 3.41$	0.23	0.09	0.04, 0.41	$F_{(1, 3471)} = 4.86^*$
{PV, RP} vs. AV	0.20	0.10	0.01, 0.39	$F_{(2, 1257)} = 4.72^{**}$	0.21	0.10	<0.01, 0.41	$F_{(2, 686)} = 3.18^*$
PV vs. RP	0.10	0.08	-0.06, 0.26		0.11	0.12	-0.12, 0.34	
Time $\times$ {PV, RP} vs. AV	0.57	0.26	0.07, 1.08	$F_{(2, 2443)} = 3.18^*$	0.43	0.26	-0.08, 0.93	$F_{(2, 1581)} = 1.84$
Time $\times$ PV vs. RP	0.28	0.12	-0.17, 0.74		0.26	0.27	-0.27, 0.78	
<i>NP2</i>								
Intercept	-0.22	0.05	-0.32, -0.13		-0.22	0.08	-0.39, -0.06	
Time	-0.59	0.10	-0.79, -0.39	$F_{(1, 4618)} = 61.64^{***}$	-0.62	0.13	-0.87, -0.36	$F_{(1, 3317)} = 9.91^{**}$
{PV, RP} vs. AV	-0.13	0.16	-0.44, 0.17	$F_{(2, 517)} = 0.99$	0.02	0.14	-0.26, 0.30	$F_{(2, 1011)} = 8.48^{***}$
PV vs. RP	-0.06	0.10	-0.24, 0.13		0.01	0.14	-0.25, 0.28	
Time $\times$ {PV, RP} vs. AV	-2.22	0.22	-2.65, -1.79	$F_{(2, 4010)} = 51.58^{***}$	-2.05	0.32	-2.67, -1.43	$F_{(2, 1898)} = 23.79^{***}$
Time $\times$ PV vs. RP	-0.55	0.22	-0.99, -0.11		-0.40	0.28	-0.94, 0.14	

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

Table 3 shows the results of the quasi-logistic linear mixed effects regression models for fixations to the patient in the three analysis time windows. During the *Verb + Adverb* region, none of the predictors reaches statistical significance, indicating that listeners' fixations to the patient did not differ between sentence types and did not change while hearing the verb and the adverb.

During the *NP1* region, there were more patient fixations in sentences with final agent pivots (1a) in which the patient was mentioned during that region. Listeners started to direct their gaze to the patient in this sentence type only toward the end of the time window which might explain that a main effect of sentence type but no interaction with time was found. There were no differences in patient fixations between sentences with sentence-final patient pivots (1b) and recent perfective marking (2) for which the agent was mentioned during this time window.

Finally, during the *NP2* region, there was a steep increase of patient fixations in sentences in which it was mentioned during this time window, i.e., patient voice and recent perfective sentences. Patient fixations decreased in sentences with agent pivots as they were mentioned sentence-finally. Additionally, in

the by-subjects analysis, there was a steeper increase of patient fixations in sentences where it was the pivot (1b). This effect is, however, barely detectable in the by-items analysis.

To test when listeners began to direct their gaze from the referent of *NP1* to the referent of *NP2*, breakpoint analyses were performed over the corresponding analysis time windows. These analyses test for discontinuities in the linear relations (Baayen, 2008), i.e., changes of direction of the regression lines for agent and patient fixations. Participants' agent fixations began to change before the beginning of *NP2* in all three sentence types (agent pivot sentences: before the first bin of *NP2* by-subjects and by-items; patient pivot sentences: before the last time bin of *NP1* by-subjects and before the first time bin of *NP2* by-items; recent perfective sentences: before the first time bin of *NP2* by-subjects and before the last time bin of *NP1* by-items). Participants' patient fixations began to change with very similar timing (agent pivot sentences: before the first bin of *NP2* by-subjects and by-items; patient pivot sentences: before the first bin of *NP2* by-subjects and by-items; recent perfective sentences: before the last time bin of *NP1* by-subjects and by-items).



**TABLE 3 | Quasi-logistic linear mixed effects regression results predicting empirical logits of fixations to the patient referent in three different sentence types.**

	By subject				By item			
	$\hat{\beta}$	SE	95% Wald CI	F statistic	$\hat{\beta}$	SE	95% Wald CI	F statistic
<i>Verb + Adverb</i>								
Intercept	-0.75	0.04	-0.84, -0.66		-0.81	0.09	-0.98, -0.64	
Time	-0.02	0.11	-0.24, 0.20	$F_{(1, 10192)} = 0.07$	0.01	0.17	-0.32, 0.34	$F_{(1, 3736)} = 0.01$
{PV, RP} vs. AV	-0.07	0.07	-0.20, 0.07	$F_{(2, 14425)} = 0.65$	-0.01	0.09	-0.19, 0.17	$F_{(2, 14124)} = 0.15$
PV vs. RP	0.04	0.08	-0.11, 0.19		0.03	0.06	-0.09, 0.16	
Time × {PV, RP} vs. AV	-0.23	0.23	-0.68, 0.22	$F_{(2, 16429)} = 0.59$	<0.01	0.20	-0.40, 0.40	$F_{(2, 27222)} = 0.02$
Time × PV vs. RP	-0.13	0.24	-0.59, 0.34		-0.05	0.24	-0.53, 0.42	
<i>NP1</i>								
Intercept	-0.70	0.04	-0.79, -0.62		-0.78	0.07	-0.92, -0.64	
Time	0.10	0.10	-0.11, 0.31	$F_{(1, 5489)} = 0.31$	0.04	0.10	-0.16, 0.24	$F_{(1, 3690)} = 0.10$
{PV, RP} vs. AV	-0.34	0.08	-0.49, -0.19	$F_{(2, 2321)} = 10.66^{***}$	-0.23	0.10	-0.43, -0.02	$F_{(2, 750)} = 3.89^*$
PV vs. RP	-0.15	0.09	-0.33, 0.02		-0.12	0.13	-0.37, 0.13	
Time × {PV, RP} vs. AV	-0.32	0.28	-0.88, 0.23	$F_{(2, 2395)} = 0.83$	-0.17	0.29	-0.74, 0.40	$F_{(2, 2015)} = 0.20$
Time × PV vs. RP	-0.12	0.19	-0.50, 0.26		-0.09	0.26	-0.60, 0.43	
<i>NP2</i>								
Intercept	-0.35	0.04	-0.44, -0.26		-0.39	0.08	-0.56, -0.23	
Time	0.56	0.12	0.33, 0.79	$F_{(1, 2228)} = 43.27^{***}$	0.55	0.14	0.27, 0.83	$F_{(1, 2624)} = 31.00^{***}$
{PV, RP} vs. AV	0.04	0.16	-0.27, 0.35	$F_{(2, 619)} = 0.28$	-0.02	0.14	-0.30, 0.27	$F_{(2, 849)} = 2.97$
PV vs. RP	-0.02	0.10	-0.22, 0.19		-0.04	0.11	-0.26, 0.17	
Time × {PV, RP} vs. AV	2.10	0.21	1.70, 2.50	$F_{(2, 10989)} = 53.99^{***}$	1.97	0.40	1.18, 2.76	$F_{(2, 1458)} = 18.13^{***}$
Time × PV vs. RP	0.47	0.18	0.11, 0.83		0.51	0.26	> -0.01, 1.03	

\* $p < 0.05$  and \*\*\* $p < 0.001$ .

In other words, before the onset of the second argument, listeners' fixations to the agent increased in agent voice-marked sentences where it was in sentence-final position and decreased in patient voice and recent perfective-marked sentences where the patient was in sentence-final position. Similarly, before the onset of NP2, patient fixations began to increase in the latter sentence types and began to decrease in sentences with agent pivots.

When controlling for agent or patient animacy (humans and animals vs. inanimates) or position within the experiment (first vs. second half), or when only items that occurred in all three conditions are included (i.e., excluding scenarios with human patients as sentences with agent pivots are prohibited in these configurations), a similar pattern of results emerges for all three analysis time windows. However, the different slopes for sentences with patient pivots and recent perfective sentences during the NP2 region that were found in the by-subjects analyses for agent and patient fixations are not consistently found when these control variables were included.

Especially the similar pattern of results that was found when the position of trials in the experiment was controlled (first vs. second half) suggests that participants' behavior was not

influenced by an expectation to encounter pronominalized or zero anaphora arguments (cf. Kroeger, 1993; Himmelmann, 1999). Participants seemed to be primed to encounter sentences with two full NP arguments by the practice trials at the beginning of the experiment; otherwise, some habituation over the course of the experiment modulating the effects of interest would have been expected.

Anticipatory baseline effects (Barr et al., 2011) influencing the interaction of time and sentence type are also not detectable when comparing the likelihood of agent or patient fixations during the preview and during the *Verb + Adverb* region (-400-200 ms relative to verb onset vs. 200 ms—NP1 onset).

## 2.4. Discussion

The results of the current visual world experiment on Tagalog suggest that listeners used the lexical semantics of the verb to determine agent and patient referents. They directed their gaze toward the agent after they heard and recognized the verb. Interestingly, listeners focused on the agent in all three sentence types, irrespective of whether it was the pivot or not and therefore also irrespective of whether it could be expected to immediately

follow the adverb or not. In contrast, while hearing the verb and the adverb, listeners did not direct their attention toward the patient.

Listeners did not seem to use information provided by the verbal morphology from which the syntactic function and the canonical position of arguments could be inferred for anticipation upon having heard the verb. If there were anticipation processes during the *Verb + Adverb* region based on syntactic information, i.e., if listeners either anticipated the final pivot argument or the linearly first NP, differences between sentences with agent pivots and sentences with patient pivots or recent perfective marking should have been found. Specifically, an increase in patient fixations would have been expected in sentences with agent pivots if anticipation was based on the linear order of NPs because in these sentences the patient canonically precedes the agent. Conversely, if anticipation was based on pivot status, an increase in patient fixation for sentences with patient pivots would have been expected. Yet, only fixations to the agent increased after listeners encountered the initial verb in all three sentence types.

Only after the adverb—during the *NP1* and *NP2* regions—did listeners gaze at agent and patient referents in their linear order. At least for the second argument (*NP2*), listeners seemed to anticipate the respective referent by directing their gaze toward the corresponding element before it was mentioned. Information provided by the verb and the first NP were integrated to predict the referent of the final argument. This interpretation is based on the consideration that programming a saccade typically takes approximately 200 ms (Duchowski, 2007) and there is also a lag between eye movements and the linguistic input of about the same time (Allopenna et al., 1998). Given that the slope of agent and patient fixations changed direction before the onset of the *NP2* region in most cases, it may be assumed that listeners programmed their eye movements toward the agent (1a) or the patient referent (1b and 2) already well before having heard and parsed the corresponding noun in the linguistic input.

The results of the current experiment thus indicate that early anticipation of arguments in Tagalog is based on semantic roles and that the agent of the event in particular attracted listeners' attention once enough information about the event had accumulated to allow the identification of agent and patient referents. In Tagalog, the possibilities for prediction upon encountering the verb are not already narrowed down by previous linguistic input, unlike in subject-initial languages where one of the verb's arguments, often the agent, has already been mentioned. Thus, in this verb-initial language, it appears that what is targeted by anticipatory processes is primarily the semantics of the event.

Altmann and Kamide (2007) argue for a linking hypothesis between language processing and eye movements that allows verbs to drive anticipatory eye movements based on the affordances of the linguistic input and the visual display (cf. also Tanenhaus et al., 2000). These affordances are the “properties of the possible interactions [...] the depicted referents] could [...] engage in” (Altmann and Kamide, 2007, p. 513). Accordingly, the presence of a frog and a fly together with the auditory presentation of “eat” conspire to create a

representation of the event that makes the frog a potential agent and the fly a potential patient. It is this episodic fit between the semantics of the described event and the depicted referents that drives listeners' eye movements toward the agent upon having heard the verb and before the agent NP was encountered.

### 3. CONCLUSIONS

A visual world experiment on a verb-initial language was presented that was set out to test what kind of information listeners are sensitive to during anticipatory processing in language comprehension. It was found that in Tagalog, listeners focus on the agent of the event upon having heard the sentence-initial verb. The lexical semantics of the verb together with the visual display allowed them to rapidly identify agent and patient referents. It seems that listeners did not use information provided by voice marking to specifically predict the syntactic functions or the linear order of arguments right after having heard the verb.

However, later in the sentence, specifically before the second noun was encountered, listeners did integrate all available information to anticipate the corresponding referent in the sentence-final position. This finding is similar to what has been found in English (Altmann and Kamide, 1999), German (Knoeferle et al., 2005) and Japanese (Kamide et al., 2003a). Thus, users of verb-initial languages also exhibit anticipatory gazes based on the linear order of arguments. Prediction of the final NP operates on a temporally more local level and occurs right before it is encountered whereas agent anticipation after the verb is independent of its position in the sentence.

It may be concluded that there are two kinds of anticipatory processes in Tagalog: one is oriented toward the sentence-level which uses verbal semantics to identify and focus on the agent of the event, the other one operates on a local scale and integrates information from the verb and the first argument to anticipate the sentence-final argument. Anticipation of the syntactic object in subject-initial languages could then possibly be seen as an instance of the latter, temporally more local, type.

Altmann and Kamide (2007) argue that anticipatory eye movements in sentence comprehension are driven by overlapping activations between representations of the visually presented objects and conceptual representations induced by the linguistic input. The results from the current experiment suggest that verbs especially facilitate anticipation based on semantic roles. Verbs provide event semantics to which potential referents in the visual display can be associated based on their affordances. Anticipatory eye movements might reflect listeners' knowledge about the dynamics of events in the world and are therefore not only reflecting “unfolding language [...] but] an unfolding (mental) world” (Altmann and Kamide, 2007, p. 515).

One possible interpretation of the findings from Tagalog is thus that language users may engage in simulation-based anticipation when processing verb-initial sentences. Huettig (2016) suggests that there are several anticipatory mechanisms in language comprehension. One of these mechanisms engages perceptual simulation of events in order to predict their outcome and the linguistic structure with which they will be represented.

Moulton and Kosslyn (2009) argue that simulation and mental imagery play a vital role for the prediction of future states of the world. Cohn and Paczynski (2013) propose an agent saliency principle that renders agents more prominent than patients in the processing of events in general (cf. also Kemmerer, 2012; Bornkessel-Schlesewsky and Schlewsky, 2013b). Upon having heard the sentence-initial verb, Tagalog listeners identified the agent referent and might have focused on it because it was the initiator of the described event and was therefore necessary to build an event structural representation and to form expectations about the remainder of the sentence. The results of the current experiment are consistent with the idea that Tagalog listeners mentally simulated the event described by the verb after having encountered it (Pulvermüller, 2005). Agents might attract the most attention during the mental simulation of events because they function as cognitive attractors as they are the instigators of these events (Bornkessel-Schlesewsky and Schlewsky, 2013a) and because the representation of agents and their actions is probably evolutionary ancient as it is already present in infants (Spelke and Kinzler, 2007).

The current findings are also in accord with approaches to sentence comprehension that assume agent identification to be an early processing step. Bornkessel and Schlewsky (2006) posit that listeners try to identify the agent as quickly as possible. Many studies also show that sentences in which the agent precedes the patient are easier to process (Schriefers et al., 1995; Traxler et al., 2002; Ferreira, 2003; Wang et al., 2009, inter alia).

Interestingly, the prominence of the agent role in comprehension processes in Tagalog has its reflexes in grammar, too. Schachter (1995) shows that both pivots and agents are privileged in different syntactic constructions (cf. also Schachter, 1976; Foley and Van Valin, 1984). Riesberg and Primus (2015) argue that even in Tagalog's symmetrical voice system, where verbs are morphologically marked for agent as well as patient pivots, agents have a special grammatical status. For example, agents are always binders of reflexives, independently of their syntactic status (Schachter, 1977). Thus, although there is no grammatical preference for agents as pivots—and patient pivots are in fact more frequent in Tagalog texts—, agents seem to take a prominent role in both processing and grammar. This is surely to be attributed to their centrality for event cognition.

Focusing on a different kind of simulation than the mental simulation of events described above, Pickering and Garrod (2013) proposed that anticipation in language comprehension emerges through prediction by (linguistic) simulation of production processes (cf. also Pickering and Garrod, 2007; Dell and Chang, 2014). Under this view, listeners use the linguistic input that they have encountered at any given point in time to build an impoverished forward production model of what they would say if they were the speaker, just as people construct forward models of motor commands (Wolpert et al., 2003). The output of this forward production process is then matched against what was actually heard. Thus, the production system would be routinely employed during comprehension by covertly imitating the speaker's behavior in order to build expectations about the following linguistic material before it is encountered.

Based on eye tracking evidence from sentence production in Tagalog, it seems that the current experiment does not directly support this view. Sauppe et al. (2013) show that in early stages of Tagalog sentence production the pivot argument plays a prominent role—irrespective of its semantic role. In a picture description experiment, speakers preferentially fixated the character that was to become the pivot argument before uttering the sentence-initial verb in order to aid encoding the morphological marking. By contrast, the current experiment found that during sentence comprehension in the presence of visual stimuli, listeners directed their attention toward the agent irrespective of which argument was the pivot of the sentence. Taken together, these results suggest Tagalog speakers and listeners prioritize the processing of distinct kinds of information during the early stages of sentence encoding and decoding.

In other words, during early phases of sentence production Tagalog speakers focus their attention on pivot arguments. During comprehension, on the other hand, Tagalog listeners focus on the agent of the event early after having heard the sentence-initial verb. This suggests that different processes may be at play and that listeners did not immediately build a forward production model of the unfolding sentence to predict upcoming words. If this would have been the case, agent and patient fixations in sentences with agent pivots and patient pivots should have differed based on the differential semantic roles of the pivot arguments. When producing a sentence, Tagalog speakers need to choose a pivot argument and encode the relevant information in form of voice affixes on the verb and case markings on the arguments. When comprehending a sentence, language users do not have to engage in choosing a pivot argument themselves. They can thus rather concentrate on verbal semantics in order quickly build a representation of the described event.

Nevertheless, effects of agent prominence can also be detected in production processes in Tagalog as the planning of sentences with agent pivots exhibits lower cognitive load requirements than the production of sentences with patient pivots (Sauppe, submitted).

It may be noted that it can not be excluded that local thematic priming between verb and arguments had an influence on listeners' gaze behavior. Kukona et al. (2012) found that anticipatory fixations in a visual world sentence comprehension experiment on English were influenced by semantic priming from verbs when there were strong associations between the verb and its arguments (e.g., *arrest* together with *policeman* and *crook*). Most notably, upon having heard the verb, listeners looked at potential agent referents even if they were not mentioned. It is to be determined in future studies whether these results can also be explained by the relative saliency of agents in the build-up of event structural representations and in how far priming effects influence early agent fixations in Tagalog.

In general it can be concluded that the structure of the input guides the uptake and integration of visual and linguistic information. The current study shows that in addition to selectional restrictions and other structural information (Altmann and Kamide, 1999; Kamide et al., 2003a; Boland, 2005), the semantic roles of event participants might also be targeted by anticipation processes. Verb-initial languages might even favor

the anticipation of semantic roles because information about the event is presented at the very beginning of an unfolding sentence and neither agent nor patient role are already (lexically) filled upon encountering the verb.

Altmann and Kamide (1999) propose that any information available to the listener is used to anticipate upcoming elements of an unfolding sentence. The results of the current experiment on Tagalog comprehension support this view. As soon as relevant information was available, listeners used selectional restrictions to identify the verb's arguments. Later on, accrued information about the event and the already encountered words was used to anticipate the final noun phrase of a sentence. Interestingly, upon having heard the verb, language users first directed their attention toward the agent, the instigator of the described event, independently of its syntactic status and its position in the sentence.

Going beyond the findings of previous visual world studies on subject-initial languages, the current experiment employed constructions in which the influence of event semantic information and syntactic information could be dissociated. It was shown that it was semantic information that was targeted early by predictive processes although syntactic information was also prominent and became relevant later. During the comprehension of languages with subject-initial word order, predictive processes on the basis of semantic roles might also operate. As mentioned in the introduction, when anticipating sentence-final syntactic objects, listeners could specifically predict the patient referent based on its role in the event described by the verb (cf. Kukona et al., 2012). This, however can not be observed as directly as in verb-initial languages because for the anticipation of sentence-final objects, semantic and syntactic information cannot be disentangled.

Tagalog has a relatively simple verbal morphology in the sense that only the semantic role of one of the arguments is cross-referenced. Future research should address whether the richness of verbal morphology has an influence on anticipatory processes. It could be possible that, e.g., person or number marking of pivot

and non-pivot arguments (or subject and object for this purpose) on an initial verb triggers different anticipatory processes because more grammatical information about arguments is provided early.

To date, there are only few studies on online language processing in verb-initial languages (most notably Sauppe et al., 2013; Norcliffe et al., 2015b; Wagers et al., 2015). These languages provide valuable means to put to test processing theories and hypotheses that were developed based on the small set of languages that is usually used in psycholinguistics (such as English, German, Dutch or Japanese; cf. Jaeger and Norcliffe, 2009 on the most studied languages in sentence production research). Making use of the grammatical diversity of the world's languages will help to refine psycholinguistic theories and to uncover processes that cannot be observed by experimentation on the "usual suspect" languages (Levinson, 2012; Norcliffe et al., 2015a).

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

This research was funded by a doctoral fellowship to the author from the International Max Planck Research School for Languages Sciences, Nijmegen, and by the Language and Cognition Department of the Max Planck Institute for Psycholinguistics, Nijmegen. The author thanks Elisabeth Norcliffe, Gabriela Garrido Rodriguez, Robert D. Van Valin, Jr, Stephen C. Levinson, Anja Latrouite and the two reviewers for discussions and helpful comments on the manuscript, and Inger M. Montemor, Philip A. Rentillo and Jem Javier for help with the stimuli creation, Ronald Fischer for technical support, and Aldrin P. Lee for making it possible to conduct the experiment at UP Diliman.

## REFERENCES

- Adams, K. L., and Manaster-Ramer, A. (1988). Some questions of topic/focus choice in Tagalog. *Ocean. Linguist.* 27, 79–101. doi: 10.2307/3623150
- Aissen, J. (1999). Markedness and subject choice in optimality theory. *Nat. Lang. Linguist. Theory* 17, 673–711. doi: 10.1023/A:1006335629372
- Alloppenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558
- Altmann, G. T. M., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Altmann, G. T. M., and Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: linking anticipatory (and other) eye movements to linguistic processing. *J. Mem. Lang.* 57, 502–518. doi: 10.1016/j.jml.2006.12.004
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511801686
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *J. Mem. Lang.* 59, 457–474. doi: 10.1016/j.jml.2007.09.002
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Front. Psychol.* 4:328. doi: 10.3389/fpsyg.2013.00328
- Barr, D. J., Gann, T. M., and Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychol.* 137, 201–207. doi: 10.1016/j.actpsy.2010.09.011
- 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
- Boland, J. E. (2005). Visual arguments. *Cognition* 95, 237–274. doi: 10.1016/j.cognition.2004.01.008
- Bornkessel, I., and Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* 113, 787–821. doi: 10.1037/0033-295X.113.4.787
- Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2013a). "Neurotypology: modeling crosslinguistic similarities and differences in the neurocognition

- of language comprehension,” in *Language Down the Garden Path: The Cognitive and Biological Basis for Linguistic Structures*, Chapter 11, eds M. Sanz, I. Laka, and M. K. Tanenhaus (Oxford: Oxford University Press), 241–252.
- Bornkessel-Schlesewsky, I., and Schlewsky, M. (2013b). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang.* 125, 60–76. doi: 10.1016/j.bandl.2013.01.010
- Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010.00025
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Cohn, N., and Paczynski, M. (2013). Prediction, events, and the advantage of Agents: the processing of semantic roles in visual narrative. *Cogn. Psychol.* 67, 73–97. doi: 10.1016/j.cogpsych.2013.07.002
- Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. B* 369:20120394. doi: 10.1098/rstb.2012.0394
- DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504
- Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice, 2nd Edn.* London: Springer.
- Ehrlich, S. F., and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *J. Verb. Learn. Verb. Behav.* 20, 641–655. doi: 10.1016/S0022-5371(81)90220-6
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cogn. Psychol.* 47, 164–203. doi: 10.1016/S0010-0285(03)00005-7
- Foley, W. A. (2008). “The place of Philippine languages in a typology of voice systems,” in *Voice and Grammatical Relations in Austronesian Languages*, Chapter 2, eds P. K. Austin and S. Musgrave (Stanford, CA: CSLI Publications), 22–44.
- Foley, W. A., and Van Valin, R. D. (1984). *Functional Syntax and Universal Grammar.* Cambridge: Cambridge University Press.
- Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression, 2nd Edn.* Thousand Oaks, CA: Sage. R package version 2.0-25.
- 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. Soft.* 59, 1–32. doi: 10.18637/jss.v059.i09
- Himmelman, N. P. (1999). The lack of zero anaphora and incipient person marking in Tagalog. *Ocean. Linguist.* 38, 231–269. doi: 10.1353/ol.1999.0010
- Himmelman, N. P. (2005). “Tagalog,” in *The Austronesian languages of Asia and Madagascar*, Chapter 12, eds A. Adelaar and N. P. Himmelman (Oxon: Routledge), 350–376.
- Himmelman, N. P. (2008). “Lexical categories and voice in Tagalog,” in *Voice and Grammatical Relations in Austronesian Languages*, Chapter 10, eds P. K. Austin and S. Musgrave (Stanford, CA: CSLI Publications), 247–293.
- Huetting, F. (2016). Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135. doi: 10.1016/j.brainres.2015.02.014
- Huetting, F., Rommers, J., and Meyer, A. S. (2011a). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol.* 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003
- Huetting, F., Singh, N., and Mishra, R. K. (2011b). Language-mediated visual orienting behavior in low and high literates. *Front. Psychol.* 2:285. doi: 10.3389/fpsyg.2011.00285
- Jaeger, T. F., and Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Lang. Linguist. Comp.* 3, 866–887. doi: 10.1111/j.1749-818X.2009.00147.x
- Kamide, Y., Altmann, G. T. M., and Haywood, S. L. (2003a). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8
- Kamide, Y., Scheepers, C., and Altmann, G. T. M. (2003b). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. *J. Psycholinguist. Res.* 32, 37–55. doi: 10.1023/A:1021933015362
- Kemmerer, D. (2012). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca’s area. *Lang. Linguist. Comp.* 6, 50–66. doi: 10.1002/lnc3.322
- Kenward, M. G., and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53, 983–997. doi: 10.2307/2533558
- Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., and Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nat. Neurosci.* 7, 1299–1301. doi: 10.1038/nn1355
- Knoeferle, P., Crocker, M. W., Scheepers, C., and Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition* 95, 95–127. doi: 10.1016/j.cognition.2004.03.002
- Kroeger, P. (1993). *Phrase Structure and Grammatical Relations in Tagalog.* Stanford, CA: CSLI Publications.
- Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., and Magnuson, J. S. (2012). The time course of anticipatory constraint integration. *Cognition* 119, 23–42. doi: 10.1016/j.cognition.2010.12.002
- Latrouite, A. (2011). *Voice and Case in Tagalog: The Coding of Prominence and Orientation.* PhD thesis, Heinrich-Heine-Universität Düsseldorf.
- Latrouite, A. (2015). “Shifting perspectives: case marking restrictions and the syntax-semantics-pragmatics interface,” in *Exploring the Syntax-Semantics-Pragmatics Interface*, eds J. Fleischhauer, A. Latrouite, and R. Osswald (Düsseldorf: Düsseldorf University Press), 285–314.
- Levinson, S. C. (2012). The original sin of cognitive science. *Top. Cogn. Sci.* 4, 396–403. doi: 10.1111/j.1756-8765.2012.01195.x
- Lockwood, H. T., and Macaulay, M. (2012). Prominence hierarchies. *Lang. Linguist. Comp.* 6, 431–446. doi: 10.1002/lnc3.345
- Magyari, L., Bastiaansen, M., de Ruiter, J. P., and Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *J. Cogn. Neurosci.* 26, 2530–2539. doi: 10.1162/jocn\_a\_00673
- Magyari, L., and de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* 3:376. doi: 10.3389/fpsyg.2012.00376
- Mishra, R. K., Singh, N., Pandey, A., and Huetting, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability – Evidence from Indian low and high literates. *J. Eye Move. Res.* 5, 1–10. Available online at: <http://www.jemr.org/online/5/1/3>
- Moulton, S. T., and Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philos. Trans. R. Soc. B* 364, 1273–1280. doi: 10.1098/rstb.2008.0314
- Norcliffe, E., Harris, A. C., and Jaeger, T. F. (2015a). Cross-linguistic psycholinguistics and its critical role in theory development: early beginnings and recent advances. *Lang. Cogn. Neurosci.* 30, 1009–1032. doi: 10.1080/23273798.2015.1080373
- Norcliffe, E., Konopka, A. E., Brown, P., and Levinson, S. C. (2015b). Word order affects the time course of sentence formulation in Tzeltal. *Lang. Cogn. Neurosci.* 30, 1187–1208. doi: 10.1080/23273798.2015.1006238
- Pickering, M. J., and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110. doi: 10.1016/j.tics.2006.12.002
- Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495
- Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effect Models in S and S-PLUS.* New York, NY: Springer. doi: 10.1007/978-1-4419-0318-1
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706
- R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.
- Riesberg, S. (2014). *Symmetrical Voice and Linking in Western Austronesian Languages.* Boston; Berlin: De Gruyter Mouton. doi: 10.1515/9781614518716
- Riesberg, S., and Primus, B. (2015). Agent prominence in symmetrical voice languages. *STUF Lang. Typol. Univ.* 68, 551–564. doi: 10.1515/stuf-2015-0023
- Robertson, S. S., and Suci, G. J. (1980). Event perception by children in the early stages of language production. *Child Dev.* 51, 89–96. doi: 10.2307/1129594
- Sauppe, S., Norcliffe, E. J., Konopka, A. E., Van Valin, R. D., and Levinson, S. C. (2013). “Dependencies first: eye tracking evidence from sentence production in Tagalog,” in *Proceedings of the 35th Annual Conference of the Cognitive Science*

- Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 1265–1270.
- Schachter, P. (1976). “The subject in Philippine languages: Topic, actor, actor-topic, or none of the above,” in *Subject and Topic*, Chapter 16, ed C. N. Li (New York, NY: Academic Press), 491–518.
- Schachter, P. (1977). “Reference-related and role-related properties of subjects,” in *Grammatical Relations*, eds P. Cole and J. M. Sadock (New York, NY: Academic Press), 279–306.
- Schachter, P. (1995). *The subject in Tagalog: Still None of the Above, Volume 15 of UCLA Occasional Papers in Linguistics*. Los Angeles, CA: Department of Linguistics, University of California.
- Schachter, P., and Otones, F. T. (1972). *Tagalog Reference Grammar*. Berkeley: University of California Press.
- Schriefers, H., Friederici, A. D., and Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *J. Mem. Lang.* 34, 499–520. doi: 10.1006/jmla.1995.1023
- Sebanz, N., and Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Top. Cogn. Sci.* 1, 353–367. doi: 10.1111/j.1756-8765.2009.01024.x
- Spelke, E. S., and Kinzler, K. D. (2007). Core knowledge. *Dev. Sci.* 10, 89–96. doi: 10.1111/j.1467-7687.2007.00569.x
- Tanenhaus, M. K., Magnuson, J. S., Dahan, D., and Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: evaluating a linking hypothesis between fixations and linguistic processing. *J. Psycholinguist. Res.* 29, 557–580. doi: 10.1023/A:1026464108329
- Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: evidence from eye movements. *J. Mem. Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836
- Wagers, M., Borja, M. F., and Chung, S. (2015). The real-time comprehension of WH-dependencies in a WH-agreement language. *Language* 91, 109–144. doi: 10.1353/lan.2015.0001
- Wang, L., Schlesewsky, M., Bickel, B., and Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the ‘subject’-preference: evidence from the online comprehension of simple sentences in Mandarin Chinese. *Lang. Cogn. Process.* 24, 1180–1226. doi: 10.1080/01690960802159937
- Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. B* 358, 593–602. doi: 10.1098/rstb.2002.1238

**Conflict of Interest Statement:** The author 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 © 2016 Sauppe. 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) or licensor 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.



# Structural Priming and Frequency Effects Interact in Chinese Sentence Comprehension

Hang Wei<sup>1</sup>, Yanping Dong<sup>2\*</sup>, Julie E. Boland<sup>3</sup> and Fang Yuan<sup>1</sup>

<sup>1</sup> School of Foreign Studies, Xi'an Jiaotong University, Xi'an, China, <sup>2</sup> Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China, <sup>3</sup> Department of Psychology, University of Michigan, Ann Arbor, MI, USA

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Lisa Lai-Shen Cheng,  
Leiden University, Netherlands  
Bing Sun,  
South China Normal University, China

### \*Correspondence:

Yanping Dong  
ypdong@mail.gdufs.edu.cn

### Specialty section:

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

**Received:** 30 June 2015

**Accepted:** 11 January 2016

**Published:** 02 February 2016

### Citation:

Wei H, Dong Y, Boland JE and Yuan F  
(2016) Structural Priming and  
Frequency Effects Interact in Chinese  
Sentence Comprehension.  
*Front. Psychol.* 7:45.  
doi: 10.3389/fpsyg.2016.00045

Previous research in several European languages has shown that the language processing system is sensitive to both structural frequency and structural priming effects. However, it is currently not clear whether these two types of effects interact during online sentence comprehension, especially for languages that do not have morphological markings. To explore this issue, the present study investigated the possible interplay between structural priming and frequency effects for sentences containing the Chinese ambiguous construction V NP1 *de* NP2 in a self-paced reading experiment. The sentences were disambiguated to either the more frequent/preferred NP structure or the less frequent VP structure. Each target sentence was preceded by a prime sentence of three possible types: NP primes, VP primes, and neutral primes. When the ambiguous construction V NP1 *de* NP2 was disambiguated to the dispreferred VP structure, participants experienced more processing difficulty following an NP prime relative to following a VP prime or a neutral baseline. When the ambiguity was resolved to the preferred NP structure, prime type had no effect. These results suggest that structural priming in comprehension is modulated by the baseline frequency of alternative structures, with the less frequent structure being more subject to structural priming effects. These results are discussed in the context of the error-based, implicit learning account of structural priming.

**Keywords:** structural priming, baseline frequency, sentence comprehension, Mandarin Chinese, inverse preference effects

## INTRODUCTION

The resolution of syntactic ambiguities during online sentence comprehension has been heavily scrutinized for it can provide evidence concerning how people draw on various sources of information. Over the years, researchers have found that factors such as structural simplicity (Rayner et al., 1983), semantic plausibility (Trueswell et al., 1994), and discourse context (Altmann and Steedman, 1988) affect syntactic ambiguity resolution. Another important source of information appears to be the baseline frequency of alternative analyses (Mitchell et al., 1995).

People seem to use baseline frequency information during online sentence processing, which might explain why different languages have different relative clause attachment preferences for sentences such as (1).

- (1) Someone stabbed the wife of the football star who was outside the house.

In sentence (1), the relative clause *who was outside the house* may be attached either to *the wife* (high in the tree structure) or to *the football star* (low attachment). Previous studies have shown a preference for high attachment in Spanish, Dutch, and French (Carreiras and Clifton, 1993, 1999; Brysbaert and Mitchell, 1996; Zagar et al., 1997; Traxler et al., 1998), whereas English shows either a weak low-attachment preference or no clear preference (Carreiras and Clifton, 1993, 1999; Traxler et al., 1998). Mitchell et al. (1995) argued that the observed cross-linguistic differences in relative clause attachment can be predicted by the relative frequency of high and low attachment in each language<sup>1</sup>.

These frequency effects represent a long-lasting effect of repeated exposure to syntactic structures in real-life settings, but there is also evidence of repeated exposure effects in experimental settings. Some early studies showed that auditory presentation of many sentences of a particular syntactic structure facilitated processing of subsequent sentences with the same structure (Mehler and Carey, 1967) or affected the interpretation of ambiguous sentences (Carey et al., 1970). More recent work found that repeated exposure to sentences containing a novel or infrequent construction could facilitate comprehension (Kaschak and Glenberg, 2004; Wells et al., 2009) or improve grammaticality ratings (Luka and Barsalou, 2005).

Moreover, there is evidence that recent exposure to as little as one instance of a structure can affect subsequent processing. Such structural priming effects were first observed in English language production (Bock, 1986), but have also been found in comprehension, in multiple languages, and even between languages in bilinguals (e.g., Hartsuiker et al., 2004; Bernolet et al., 2007, 2009). One line of research shows that priming in comprehension is lexically-dependent in that structural priming occurs when prime and target trials have certain lexical overlap but not otherwise (e.g., Branigan et al., 2005; Arai et al., 2007; Traxler and Tooley, 2008; Tooley et al., 2009; Chen, 2010). At the same time, there is evidence that priming in comprehension may occur without verb repetition (Scheepers and Crocker, 2004; Traxler, 2008; Thothathiri and Snedeker, 2008a,b).

More recently, there has been growing evidence of lexically-independent priming effects, both in processing parallel constructions (e.g., *A demanding boss said that a lazy worker did not do the job properly*; Sturt et al., 2010) and in comprehending dative constructions that contain verb anomalies (e.g., *The waitress brunks/exists the book to the monk*; Ivanova et al., 2012). In fact, lexically-independent priming in comprehension appears to have the same strength as priming in production when both were examined within a single experiment that

<sup>1</sup>The effects of baseline frequency upon Chinese relative clause processing appear to be controversial. Although subject relative clauses (SRCs, e.g., raokai baoan de jizhe “the reporter that bypassed the guard”) are more frequent than object relative clauses (ORCs, e.g., baoan raokai de jizhe “the reporter that the guard bypassed”) in Mandarin Chinese, prior research has led to conflicting results regarding ease of processing SRCs vs. ORCs (e.g., Hsiao and Gibson, 2003; Lin and Bever, 2006; Chen et al., 2008; Wu et al., 2012; Gibson and Wu, 2013; Wang and Bing, 2013). We will return to this issue in Discussion.

involved identical materials and participants (Tooley and Bock, 2014). Finally, by employing more participants and new analyses methods, Pickering et al. (2013) found both lexically dependent and lexically independent priming effects in comprehending prepositional phrase attachment ambiguities (cf. Branigan et al., 2005). More importantly, these effects persisted, in that they were unaffected by whether prime and target sentences were adjacent or separated by one or two fillers.

The existence of lexically specific and lexically independent priming effects during comprehension are consistent with previous sentence processing research demonstrating frequency-based preferences that are associated with both specific lexical items (Trueswell et al., 1993) and syntactic constructions (Brysbaert and Mitchell, 1996). This consistency in findings suggests that structural priming and frequency-based syntactic preferences may have a common source. Both types of effects stem from experience with language (though at different timescales). Frequency-based preferences have presumably been acquired over a lifetime experience with the language, whereas structural priming effects result from very recent exposure to a single instance of a given structure.

Structural priming effects may interact with frequency-based syntactic preferences, as suggested by several language production studies. For example, a number of Dutch and German studies reported that priming in production may exhibit an *inverse frequency/preference effect*. This refers to the fact that structures that are produced less often seem to exhibit greater priming effects and vice versa (Hartsuiker and Kolk, 1998; Hartsuiker et al., 1999; Hartsuiker and Westenberg, 2000; Scheepers, 2003; Bernolet and Hartsuiker, 2010). The existence of such effects indicates that priming in production is modulated by the baseline frequency of alternative structures.

Recently, Segart et al. (2011) showed that the preference ratio of two syntactic alternatives is a crucial determinant of structural priming effects in Dutch language production. They measured both the proportion of passive/active picture descriptions following a passive or active prime, and the response latency of the picture description. In Experiment 1, priming with the less frequent passive structure led to an increase in passive picture descriptions. There was no corresponding increase in active picture descriptions following primes of the more frequent active structure, but response latencies were decreased following active primes. In other words, priming increased the frequency of the dispreferred alternative and decreased the response latency of the preferred alternative.

If similar mechanisms are involved in priming in comprehension and production (as suggested in cross-modal structural priming, both from production to comprehension, e.g., Branigan et al., 2005; and from comprehension to production, e.g., Bock et al., 2007), it is expected that priming in comprehension is also modulated by the baseline frequency of alternative structures. Until now, however, comparatively little is known about the possible relationship between structural priming and frequency effects during comprehension.

Although the influence of frequency upon structural priming in comprehension has not been systematically explored, incidental reports from the literature suggest



that priming in comprehension is affected by the baseline frequency of alternative analyses. In a visual-world eye-tracking experiment, Scheepers and Crocker (2004) explored whether constituent order ambiguity resolution in German was subject to priming. In German, both subject-verb-object (SVO) and object-verb-subject (OVS) orderings are permitted, though the former is more frequent. Moreover, the German case-marking system is partially ambiguous so that, sentence-initial NPs like “Die Krankenschwester . . . ” (The nurse [feminine, singular] . . . ) can be interpreted as either subject or object. Scheepers and Crocker found that the kind of constituent order being processed in a prime trial affected the constituent ordering preferences in a target trial. More crucially, relative to the neutral prime condition, only the less frequent OVS structure elicited reliable priming effects, whereas the more frequent SVO structure merely induced a numerical trend. These results suggest the possibility that priming in comprehension is constrained by the baseline frequency, but this conclusion is weakened by the fact that the relevant evidence derives from an experiment conducted for other purposes.

Additionally, note that the existing evidence comes mainly from Germanic languages (English, Dutch, German) in which syntactic functions are marked through morphological variations (in case, gender, number, etc.). The observed priming effects might be elicited by syntactic representations as well as by low-level morphosyntactic features. It is important, therefore, to explore whether structural priming exists, and more importantly, whether it interacts with frequency-based syntactic preferences in languages where syntactic functions are not marked morphologically.

In a recent study, Chen (2010) investigated the online processing of Chinese sentences containing an embedded relative clause (e.g., *guma chengzan de biao ge haizai guowai dushu* “The cousin that the aunt praised has been pursuing his study abroad”). Unlike Germanic languages, Chinese does not have morphosyntactic categories and consequently, syntactic functions are not marked morphologically in this language. This ensures that potential priming effects must rely on the preservation of abstract syntactic representations rather than low-level morphosyntactic features. Using eye-tracking and ERP, Chen found structural priming effects when the verb in the relative clause was repeated across the prime and target sentences but not otherwise (see also Chen et al., 2012, 2013). It seems, therefore, that structural priming in Chinese sentence comprehension is at least partly lexically-dependent.

More recently, Wei (2013) investigated the online processing of temporarily ambiguous Chinese sentences containing the construction V NP1 *de* NP2 (e.g., *baifang zuojia de pengyou* “visit writer *de* friend”). This construction allows for two possible structural analyses, namely, VP [<sub>VP</sub>V [<sub>NP</sub>[<sub>DeP</sub>NP1 *De*] NP2]] (to visit the writer’s friend), and NP [<sub>NP</sub>[<sub>DeP</sub> [<sub>VP</sub>V NP1] *De*] NP] (a friend who is visiting the writer). These two analyses have identical surface form but distinct underlying syntactic structures. In addition, although this construction allows for two analyses, the relative frequency of the two analyses differs. In the context of V NP1 *de* NP2, over 700 out of 1000 items

randomly selected from a corpus<sup>2</sup> were used as NP (Zhang et al., 2000).

In a self-paced reading experiment, Wei (2013) found that when the construction V NP1 *de* NP2 was disambiguated to the less frequent VP structure (the VP target condition) there was an effect of prime type: Participants experienced more processing difficulty following an NP prime relative to following either a VP prime or a neutral baseline. In contrast, when the ambiguity was resolved to the more frequent NP structure (the NP target condition), prime type had no effect. These results suggest that structural priming in Chinese sentence comprehension is modulated by frequency-based syntactic preferences, with the less frequent structure being more sensitive to structural priming effects.

However, Wei’s (2013) conclusions are weakened by several flaws in the research design. First, each prime sentence was followed by a comprehension question, which might interrupt structural priming. Second, participants were exposed to disproportionate numbers of NP or VP structures in different target conditions. In the NP target condition, there were 24 sentences of the NP structure (6 NP primes plus 18 NP-disambiguated target sentences) but there were only six sentences of the VP structure (6 VP primes). The rate of NP vs. VP structure was 4:1. The opposite holds for the VP target condition. The difference in the base rate of the two structures makes any comparisons between the two target conditions problematic: Having been exposed to an unbalanced number of NP or VP structure, participants might display different processing tendencies across conditions. They might tend to interpret the ambiguous construction V NP1 *de* NP2 as NP in the NP target condition but interpret it as VP in the VP target condition. These different processing tendencies, in turn, might be a confounding variable in the assessment of structural priming effects.

Because of the important theoretical implications of Wei (2013), we reinvestigated the priming of the Chinese ambiguous construction V NP1 *de* NP2 while avoiding these two problems in his research design. First, we removed all comprehension questions following the prime sentences. Second, we added filler sentences of the opposite structure to balance the total number of NP and VP sentences under each target condition (see details in Materials). We expected to replicate the pattern of results from Wei (2013) with the improved design, particularly the interplay between structural priming and baseline frequency in Chinese online sentence comprehension.

## MATERIALS AND METHODS

The purpose of the present experiment was to investigate whether online processing of the ambiguous construction V NP1 *de* NP2 could be affected by the prior presentation of a single prime sentence, and more crucially, whether the strength of structural priming could be modulated by the baseline frequency of alternative structures. In the present study,

<sup>2</sup>These items were selected from the Corpus for Studies of Modern Chinese (Beijing Language and Culture University, 1995), which contains 1.24 million words covering a wide range of genres.

frequency-based syntactic preference for the construction V NP1 *de* NP2 was determined on the basis of corpus data and sentence-fragment completion data reported in previous research (e.g., Zhang et al., 2000; Hsieh et al., 2009). According to a corpus analysis, the ratio of NP to VP stands at 7:3 (Zhang et al., 2000. See Note 1). The NP advantage was even more distinct in an off-line sentence fragment completion test conducted by Hsieh et al. (2009). Their data showed that the fragment *V NP1 de* was typically continued with a noun phrase, which constituted part of an NP completion 95% of the time (911/960), with VP completions accounting for 5% only. These data make it evident that the initial preference ratio is strongly biased toward the NP analysis. Additionally, in a self-paced reading experiment, Zhang et al. (2000) found that processing difficulty occurred immediately when a semantically equibaised item was disambiguated to the VP analysis but no difficulty occurred when it was disambiguated to the NP analysis, confirming that NP is the preferred analysis for the Chinese reader.

Because syntactic ambiguity resolution has been found to be affected by factors such as semantic plausibility information and discourse context (Van Gompel and Pickering, 2007), we used semantically equibaised items only, thereby holding constant any possible semantic effects. In addition, the construction V NP1 *de* NP2 was placed at the beginning of each sentence, so that its interpretation would not be affected by prior context other than the three types of prime sentences.

## Participants

Fifty-five participants from Xi'an Jiaotong University received a small payment for taking part in the experiment. One participant answered more than 20% of the comprehension questions incorrectly and was excluded from further analyses. The remaining 54 participants answered more than 80% of the comprehension questions correctly, with an average accuracy rate of 94%. Ethical approval for the experiment was granted by the School of Foreign Studies Academic Committee at Xi'an Jiaotong University.

## Materials

The experiment employed a 3 (prime type) × 2 (target type) mixed design, with six experimental lists. Prime type (NP, VP, and neutral primes) was manipulated both within-participants and within-items, and target type (NP- and VP-disambiguated) was between-participants and within-items. The 18 prime sentences and 18 semantically equibaised target sentence pairs were adopted from Wei (2013). See **Table 1** for a set of examples. A complete list of prime and target sentences appears in the Supplementary Material.

The same 18 prime sentences (6 of each type) appeared on all six experimental lists, always immediately preceding a target sentence. Six sentences started with V NP1 *de* NP2 that could only be analyzed as an NP (NP primes), six sentences started with V NP1 *de* NP2 that could only be analyzed as a VP (VP primes), and six sentences of irrelevant structures (neutral primes). The NP and VP primes were created by manipulating the thematic role of NP2. Note that the ambiguity of the construction V

**TABLE 1 | Sample prime and target sentences.**

### PRIME SENTENCES

(1) weizao zhengju de lvshi yiwei buhui youyen faxian zhenxiang (NP prime)  
falsify evidence *DE* lawyer think no person find out truth.

*The lawyer who falsified evidence thought that nobody would find it out.*

(2) shusan shanshang de youke zhiwai, tamen qidong jinji yu'an (VP prime)  
evacuate mountain *DE* tourist besides, they start emergency plan.

*Besides evacuating tourists in the mountain, they started the emergency plan.*

(3) qingwa shi liangqi dongwu, keyi zai ludi he shuizhong shenghuo (Neutral prime)  
frog is amphibian, can land and water live

*Frogs are amphibians and can live both on land and in water.*

### TARGET SENTENCES

(1) baifang zuojia de pengyou jianyi zuojia chuanguo yibu huaju  
(NP-disambiguated target)

visit writer *DE* friend suggest writer write one modern drama

*A friend who was visiting the writer advised him to write a modern drama.*

(2) baifang zuojia de pengyou qijian, XiaoGuo youle xinde xiangfa  
(VP-disambiguated target)

visit writer *DE* friend during, Guo had new idea

*During a visit to the writer's friend, Guo had a new idea.*

NP1 *de* NP2 hinges upon, among other things, the thematic role information associated with NP2. If NP2 can be an agent as well as a theme, then both NP and VP analyses are plausible. If NP2 can only be an agent or a theme, then only the NP or the VP analysis is plausible respectively. This allowed us to construct NP and VP primes by varying the thematic role associated with NP2. Neutral primes started with a subject-predicate construction, which was followed by a second clause commenting on the subject/topic (e.g., qingwa shi liangqi dongwu, keyi zai ludi he shuizhong shenghuo “*Frogs are amphibians and can live both on land and in water*”). Unlike NP and VP primes, these sentences were structurally unrelated to V NP1 *de* NP2, and were considered unlikely to trigger a particular interpretation of the target construction under investigation. Thus, they provide a baseline with which the effectiveness of NP and VP primes can be compared.

Each experimental list included either 18 NP target sentences or 18 VP target sentences. Each target sentence began with a phrase that was semantically equibaised between the NP analysis and the VP analysis. The two conditions differed in the disambiguating region, which was either resolved as an NP or as a VP. There were no lexical or semantic connections between prime and target sentences.

Note that the target sentences were always resolved as the same structure on an experimental list. To balance the total number of NP and VP structures that each participant read, we added 18 filler sentences of the VP structure (VP fillers) to the three experimental lists of the NP target condition, and 18 fillers of the NP structure (NP fillers) to the three experimental lists of the VP target condition. An NP filler started with a V-NP-*de*-NP string that was semantically biased toward the NP analysis (e.g., shuluo haizi de mama “*scold child de mother*”) and was fully disambiguated as NP by the following word (usually

a verb, as shown in 2). A VP filler started with a V-NP-*de*-NP string that was semantically biased toward the VP analysis (e.g., caifang jiaoshou de furen “interview professor *de* wife”) and was fully disambiguated as VP by the following word (usually a conjunction or preposition, as shown in 3).

- (2) shuluo haizi de mama houhui ziji guande buyan (NP filler)  
scold child *de* mother regret self discipline not strict  
*The mother who scolded the child regretted that she had been too lax.*
- (3) caifang jiaoshou de furen zhihou, jizhe xiele yipian baodao (VP filler)  
interview professor *de* wife after, journalist write LE one report  
*After interviewing the professor’s wife, the journalist wrote a report.*

In addition, there were 48 other filler sentences of various structures. None of these fillers contained the construction V NP1 *de* NP2, and they were lexically and semantically unrelated to the experimental sentences. Each participant read 102 sentences in total (18 primes, 18 targets, 18 NP/VP fillers, and 48 other fillers). The experimental sentences and fillers were presented in a single random order across lists, with the constraints that at least one filler sentence intervened between each prime-target pair, and that none of the prime-target pairs was immediately preceded by an NP or VP filler.

To encourage comprehension, some of the filler sentences were followed by a question, including the 18 VP/NP fillers as well as 24 fillers of irrelevant structures. Participants pressed the “Y” or “N” buttons to give their answers and received no feedback. Half of the questions required a “yes” response and half a “no” response.

## Procedure

The experiment was conducted using E-prime software. Participants were tested individually and randomly assigned to an experimental list. They were instructed to read the sentences at a pace that closely approximated their normal reading speed, and to read them carefully so as to answer the questions that followed some of the sentences. Each sentence was presented on a single line, beginning from the left edge of the screen. The first screen that participants saw outlined a sentence using a series of underlines, with each word being covered by a single underline. Participants would then press “Enter” on the keyboard to uncover and read the first word. With each press of the “Enter” button, the next word would be uncovered and the previous word would be covered by an underline again. The construction V NP1 *de* NP2 was presented as one region<sup>3</sup>, followed by the remaining part of the sentences that was presented word by

<sup>3</sup>The primary reason for presenting the construction V NP1 *de* NP2 as one region was to boost possible priming effects associated with the construction, which might be much weaker if presented word by word. Admittedly, this is not a natural way of reading. However, given that all sentences (including practice sentences, experimental sentences, and fillers) were presented this way, it might not create different expectations among the participants. Debriefing showed that participants were not aware of the structural features of the experimental sentences, nor were they aware of the purpose of the experiment.

word. Including a practice session of 12 sentences, the whole experiment lasted about 18 min.

## Scoring and Data Analysis

For each type of target sentence, we analyzed three regions, indicated by slashes “/” and numbers in sentences (4) and (5):

- (4) VP-disambiguated target  
baifang zuojia de pengyou /qijian<sub>1</sub>/ XiaoGuo<sub>2</sub>/ youle<sub>3</sub>/  
visit writer *DE* friend during, Guo have  
xinde xiangfa  
new idea  
*During a visit to the writer’s friend, Guo had a new idea.*
- (5) NP-disambiguated target  
baifang zuojia de pengyou/ jianyi<sub>1</sub>/ zuojia<sub>2</sub>/ chuanguozuo<sub>3</sub>/  
visit writer *DE* friend advise writer write  
yibu huaju  
one modern drama  
*A friend who was visiting the writer advised him to write a modern drama.*

The first region for the statistical analysis was the disambiguating word, which was a conjunction or a preposition in the VP target condition, and a verb in the NP target condition<sup>4</sup>. The second and the third regions corresponded to the post-disambiguation regions. The two post-disambiguation regions were included because self-paced reading task might be subject to spill-over effects, with processing load being carried over from one display to the next (Trueswell et al., 1993; Spivey-Knowlton and Sedivy, 1995). The other regions were not included for further analyses because analyzing those regions did not test any hypothesis-bearing predictions.

Prior to analyzing the data, we eliminated any reading times less than 100 milliseconds (ms) or greater than 2500 ms. This criterion eliminated 0.4% of the data. Next, any reading times that were 3 standard deviations above or below the by-participants condition mean were replaced by the cutoff values (Ratcliff, 1993), which affected 1.8% of the data.

## RESULTS

**Table 2** presents mean reading times by region and condition for the experiment. Mean reading times for the disambiguation and the two post-disambiguation regions were subject to 3 (prime type) × 2 (target type) mixed ANOVAs, with separate analyses treating participants and items as random factors. Prime type was treated as a within-participant and within-item factor, and target type was treated as a between-participant and within-item factor.

<sup>4</sup>The syntactic category difference is unavoidable: To disambiguate the construction V NP1 *de* NP2 as NP and VP respectively, we had to use words of different syntactic categories. We acknowledge that this weakens comparisons between VP and NP targets, and the possible effect of target type would be difficult to interpret for that reason (we thank an anonymous reviewer for raising this issue). However, the crucial interaction between prime and target is less problematic, because our post-hoc tests showed that it was caused by the prime type differences within each target type (see Results for details).

**TABLE 2 | Means and standard deviations in milliseconds for reading times in each region and condition.**

Target	Prime	Disambiguation Region		Post-disambiguation Region 1		Post-disambiguation Region 2	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
VP-disambiguated	Neutral	508	160	478	123	388	63
	VP	524	156	511	130	386	88
	NP	552	178	565	215	408	83
NP-disambiguated	Neutral	545	147	489	175	420	98
	VP	561	165	436	120	431	132
	NP	540	152	456	96	421	94

## Disambiguation Region

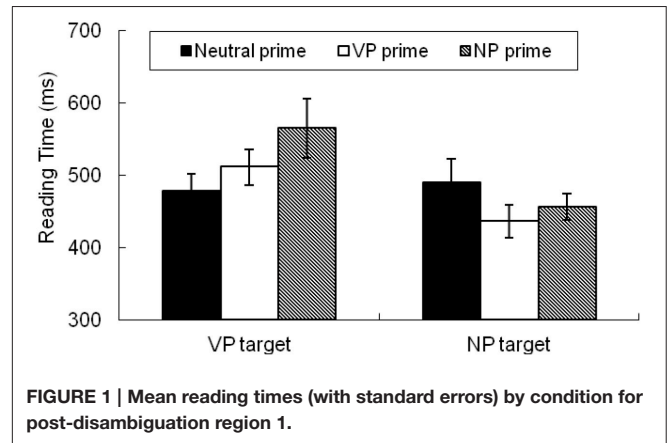
In the disambiguation region, there were no effects of prime type or target type, nor an interaction (all  $F_s < 1$ ).

## Post-Disambiguation Region 1

The interaction between prime and target types was significant both by participants and by items<sup>5</sup>:  $F_{1(2, 104)} = 4.11, p = 0.019, MSE = 12, 514$ ;  $F_{2(2, 68)} = 4.09, p = 0.021, MSE = 9285$ . The effect of prime type was not reliable [ $F_{1(1, 52)} = 1.56, p = 0.214$ ;  $F_{2(1, 34)} = 1.39, p = 0.255$ ]. The effect of target type approached significance in the by-participants analysis [ $F_{1(1, 52)} = 3.27, p = 0.077, F_{2(1, 34)} = 2.66, p = 0.112$ ], with the less frequent VP-disambiguated targets tending to be read more slowly than the more frequent NP-disambiguated targets. Tests for simple effects showed differences between prime types for the VP-disambiguated targets [ $F_{1(2, 104)} = 4.11, p = 0.019$ ;  $F_{2(2, 68)} = 3.83, p = 0.027$ ], but not for the NP-disambiguated targets [ $F_{1(2, 104)} = 1.56, p = 0.215$ ;  $F_{2(2, 68)} = 1.65, p = 0.199$ ].

**Figure 1** presents mean reading times by condition for post-disambiguation region 1, which gives a good indication of the variation in processing load associated with syntactic disambiguation as a function of prime and target types. As **Figure 1** shows, for the VP-disambiguated targets, the NP prime condition had longer reading times than the Neutral prime condition [565 vs. 478 ms,  $t_{1(26)} = 2.40, p = 0.024$ ;  $t_{2(17)} = 2.78, p = 0.013$ ]. The NP prime condition also had longer reading times than the VP prime condition, though the difference between them only approached significance in the by-participants analysis [565 vs. 511 ms,  $t_{1(26)} = 1.83, p = 0.079$ ] and was not reliable in the by-items analysis [ $t_{2(17)} = 1.22, p = 0.239$ ]. The difference between the VP prime condition and

<sup>5</sup>This pattern of results differs from that of Wei (2013), who found a significant interaction between prime and target types right in the disambiguation region (rather than in the post-disambiguation region). One possible reason for the later occurrence of the expected effects in the present study might have been the use of filler sentences of the opposite structure under the two target conditions (recall that there were 18 filler sentences of the NP structure under the VP target condition, and conversely, there were 18 fillers of the VP structure under the NP target condition. See Materials for details).

**FIGURE 1 | Mean reading times (with standard errors) by condition for post-disambiguation region 1.**

the Neutral condition was not reliable [ $511$  vs.  $478$  ms,  $t_{1(26)} = 1.41, p = 0.170$ ;  $t_{2(17)} = 1.15, p = 0.268$ ].

## Post-Disambiguation Region 2

In the second post-disambiguation region, the effect of prime type was not significant (both  $F_s < 1$ ), nor was the effect of target type [ $F_{1(1, 52)} = 2.21, p = 0.143, MSE = 16,803$ ;  $F_{2(1, 34)} = 1.91, p = 0.176, MSE = 12,521$ ], or the interaction between prime and target, both  $F_s < 1$ .

## DISCUSSION

We investigated the interaction between structural priming and frequency-based syntactic preferences in Chinese sentence comprehension in a self-paced reading experiment. We found that when the temporarily ambiguous target sentence was resolved to the less frequent analysis, participants experienced more processing difficulty (as indicated by the increase in reading times in the first post-disambiguation region) if they had just read a prime sentence of the alternative structure as opposed to a prime sentence of the same structure or a neutral baseline. When the ambiguity was resolved to the more frequent/preferred structure, prime type had no effect on the target sentence.

This pattern of results is similar to that of Wei (2013), though in the present study we eliminated the potential problems existing in Wei's research design (e.g., We removed the comprehension question following each prime sentence and presented participants with equal number of NP and VP structures across conditions. See Materials for details). The current results, paired with those of Wei (2013), suggest that priming in Chinese sentence comprehension is modulated by the baseline frequency of alternative structures, with the less frequent structure being more sensitive to structural priming effects.

The effects of baseline frequency upon structural priming during online sentence comprehension are consistent with previous comprehension studies across several languages (e.g., English, Dutch, French, Spanish), which showed that syntactic ambiguity resolution was affected by the frequency of alternative analyses, with participants preferring more frequent over less frequent ones (see Van Gompel and Pickering, 2007, for

a review). Such frequency-based syntactic preferences have presumably been acquired over a lifetime of experience with the language. There has been evidence that repeated exposure to a given structure under experimental conditions produces cumulative priming effects (Kaschak and Glenberg, 2004; Wells et al., 2009). It seems perfectly plausible that repeated exposure to syntactic structures in real-life settings could have similar effects upon the language processing system.

As noted earlier, the effects of frequency seem to be controversial in Chinese relative clause processing. Some studies showed that the more frequent subject relative clauses were easier to process than the less frequent object relative clauses (Kuo and Vasishth, 2006; Lin and Bever, 2006; Liu et al., 2011; Wu et al., 2012), while other studies yielded the opposite results (Hsiao and Gibson, 2003; Chen et al., 2008; Zhang and Yang, 2010; Zhou et al., 2010; Gibson and Wu, 2013; Wang and Bing, 2013). These conflicting results might be partly due to the experimental materials used in these studies. As Wu et al. (2012) noted, most prior studies employed relative clauses that contained two animate NPs (e.g., SRC, *raokai bao'an de jizhe* "the reporter that bypassed the guard," and ORC, *bao'an raokai de jizhe* "the reporter that the guard bypassed"). This differs from the animacy configurations found in written and spoken corpora (Pu, 2007; Wu, 2009), where the two NPs typically have contrastive animacy configurations, with subject NP being animate and object NP being inanimate (e.g., SRC, *duokai shikuai de jizhe* "the reporter that dodged the stone," and ORC, *jizhe duokai de shikuai* "the stone that the reporter dodged"). When animacy of the two NPs conformed to this pattern, subject relative clauses were found to be easier to process (Liu et al., 2011; Wu et al., 2012). It seems, therefore, that the baseline frequency also constrains Chinese relative clauses processing, though its effects are modulated by semantic information such as animacy of NPs.

The finding that the less frequent structure was more sensitive to structural priming effects is also consistent with previous research on priming in German sentence comprehension (Scheepers and Crocker, 2004). In their visual-world eye-tracking experiment, Scheepers and Crocker found that the less frequent OVS structure elicited reliable priming effects in German sentence comprehension, whereas the more frequent SVO structure merely induced a numerical but nonsignificant trend. This is similar to what we have found in the present study, though we used a different research paradigm (i.e., self-paced reading) and investigated structural priming in a topologically distinct language (in which alternative structures are not accompanied by morphological variations). Thus, the results of the present study provide cross-linguistic evidence attesting to the constraining effects of baseline frequency upon structural priming in comprehension, and suggest that such effects are not confined to a particular language or research paradigm.

The finding of inverse preference effects in Chinese sentence comprehension has important theoretical implications. There is an ongoing debate over whether structural priming effects are caused by a short-term residual activation mechanism (Pickering and Branigan, 1998), or reflects a form of long-term implicit

learning of syntactic structures (Bock and Griffin, 2000; Chang et al., 2006). The finding of such inverse preference effects in comprehension is consistent with the implicit learning account of structural priming, according to which the language processing system learns more about representations that are experienced less often (which follows from error-based learning, a strategy typically adopted in implicit learning algorithms). The presence of such effects appears to be inconsistent with the residual activation account of structural priming, according to which priming originates from transient activation that is sensitive to the immediately preceding structure but may be less affected by the baseline frequency of a given structure.

As mentioned before, the constraints of baseline frequency upon structural priming have also been observed in studies on priming in language production, where structures that are produced less often seem to be more effective in eliciting the target production (Hartsuiker and Kolk, 1998; Scheepers, 2003; Bernolet and Hartsuiker, 2010). However, the inverse preference effects in production appear to hold for response choice, but not necessarily for response time (Segaert et al., 2011; cf. Corley and Scheepers, 2002). In contrast, our inverse preference effect is a type of response time effect, as was the analogous effect observed by Scheepers and Crocker (2004) and Wei (2013). Thus, although structural priming in both comprehension and production appear to be constrained by the baseline frequency of alternative structures, the constraints might manifest differently in production and comprehension.

Note that the priming effects in the present study (as well as in Wei, 2013) occurred in the absence of lexical repetition between prime and target sentences. This stands in contrast to Chen and colleagues' studies (Chen, 2010; Chen et al., 2012, 2013) which showed that priming in Chinese sentence comprehension was crucially dependent upon verb repetition. One possible reason underlying this disparity might be due to the different structures employed in these studies. Chen and colleagues' experimental sentences involved relative clauses, whereas the present study and Wei (2013) employed sentences containing the ambiguous phrase V NP1 *de* NP2. There has been evidence that structural priming in English relative clause occurred only when the same verb was used across the prime and target sentences (Ledoux et al., 2007; Traxler and Tooley, 2008; Tooley et al., 2009), whereas priming in other syntactic structures appeared to be less dependent on lexical repetition (e.g., Traxler, 2008; Ivanova et al., 2012; Pickering et al., 2013; Tooley and Bock, 2014). More investigations should be undertaken to explore this issue.

A final point concerns the finding that relative to the neutral prime condition, VP primes produced no substantial impact on participants' processing of VP-disambiguated targets. This null effect might partly be attributed to the low baseline frequency of the VP structure (Recall that VP accounts for less than 30% in corpus data and only about 5% in sentence completion data; Zhang et al., 2000; Hsieh et al., 2009). As Segaert et al. (2011) suggested, only when the bias against the less preferred syntactic alternative is sufficiently weak a response latency effect prevails. If this reasoning is on the right track, it raises the possibility that a stronger manipulation of the prime conditions would show priming effects of the VP primes.

In sum, the present study provides cross-linguistic evidence that the strength of structural priming during online sentence comprehension is constrained by the baseline frequency of the alternative structures. Particularly, the less frequent structure seems to exhibit greater structural priming effects. The results of the present study, paired with the findings from previous research on other languages, suggest that the constraining effects of baseline frequency upon structural priming (and language comprehension in general) may reflect the operation of some general mechanism that is inherent to the human language processing system.

## REFERENCES

- Altmann, G., and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition* 30, 191–238. doi: 10.1016/0010-0277(88)90020-0
- Arai, M., Van Gompel, R. P. G., and Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cogn. Psychol.* 54, 218–250. doi: 10.1016/j.cogpsych.2006.07.001
- Bernolet, S., and Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition* 114, 455–461. doi: 10.1016/j.cognition.2009.11.005
- Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2007). Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 931–949. doi: 10.1037/0278-7393.33.5.931
- Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2009). Persistence of emphasis in language production: a cross-linguistic approach. *Cognition* 112, 300–317. doi: 10.1016/j.cognition.2009.05.013
- Bock, K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi: 10.1016/0010-0285(86)90004-6
- Bock, K., Dell, G. S., Chang, F., and Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition* 104, 437–458. doi: 10.1016/j.cognition.2006.07.003
- Bock, K., and Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *J. Exp. Psychol. Gen.* 129, 177–192. doi: 10.1037/0096-3445.129.2.177
- Branigan, H. P., Pickering, M. J., and Mclean, J. F. (2005). Priming prepositional-phrase attachment during comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 468–481. doi: 10.1037/0278-7393.31.3.468
- Brybaert, M., and Mitchell, D. C. (1996). Modifier attachment in sentence parsing: evidence from Dutch. *Q. J. Exp. Psychol.* 49A, 664–695.
- Carey, P., Mehler, J., and Bever, T. G. (1970). Judging the veracity of ambiguous sentences. *J. Verbal Learn. Verbal Behav.* 9, 243–254. doi: 10.1016/S0022-5371(70)80058-5
- Carreiras, M., and Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Lang. Speech* 36, 353–372.
- Carreiras, M., and Clifton, C. (1999). Another word on parsing relative clauses: eyetracking evidence from Spanish and English. *Mem. Cognit.* 27, 826–833. doi: 10.3758/BF03198535
- Chang, F., Dell, G. S., and Bock, K. (2006). Becoming syntactic. *Psychol. Rev.* 113, 234–272. doi: 10.1037/0033-295X.113.2.234
- Chen, B., Ning, A., Bi, H., and Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object relative clause. *Acta Psychol.* 129, 61–65. doi: 10.1016/j.actpsy.2008.04.005
- Chen, Q. (2010). *Syntactic Priming in Chinese Sentence Comprehension: Evidence from ERPs and Eyetracking*. (Unpublished doctoral dissertation). Nanjing Normal University, Nanjing.
- Chen, Q., Tan, D., and Cai, H. (2012). Syntactic priming in Chinese sentence comprehension: evidence from eye movements and ERP. *Advan. Psychol. Sci.* 20, 1727–1734. doi: 10.3724/SP.J.1042.2012.01727
- Chen, Q., Xu, X., Tan, D., Zhang, J., and Zhong, Y. (2013). Syntactic priming in Chinese sentence comprehension: evidence from Event-Related Potentials. *Brain Cogn.* 83, 142–152. doi: 10.1016/j.bandc.2013.07.005
- Corley, M., and Scheepers, C. (2002). Syntactic priming in English sentence production: categorical and latency evidence from an internet-based study. *Psychonom. Bull. Rev.* 9, 126–131. doi: 10.3758/BF03196267
- Gibson, E., and Wu, I. (2013). Processing Chinese relative clauses in context. *Lang. Cogn. Process.* 28, 125–155. doi: 10.1080/01690965.2010.536656
- Hartsuiker, R. J., and Kolk, H. H. J. (1998). Syntactic persistence in Dutch. *Lang. Speech* 41, 143–184.
- Hartsuiker, R. J., Kolk, H. H. J., and Huiskamp, P. (1999). Priming word order in sentence production. *Q. J. Exp. Psychol.* 52, 129–147.
- Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychol. Sci.* 15, 409–414. doi: 10.1111/j.0956-7976.2004.00693.x
- Hartsuiker, R. J., and Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition* 75, B27–B39. doi: 10.1016/S0010-0277(99)00080-3
- Hsiao, F., and Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition* 90, 3–27. doi: 10.1016/S0010-0277(03)00124-0
- Hsieh, Y., Boland, J., Zhang, Y., and Yan, M. (2009). Limited syntactic parallelism in Chinese ambiguity resolution. *Lang. Cogn. Process.* 24, 1227–1264. doi: 10.1080/01690960802050375
- Ivanova, I., Pickering, M. J., Branigan, H. P., Mclean, J. F., and Costa, A. (2012). The comprehension of anomalous sentences: evidence from structural priming. *Cognition* 122, 193–209. doi: 10.1016/j.cognition.2011.10.013
- Kaschak, M. P., and Glenberg, A. M. (2004). This construction needs learned. *J. Exp. Psychol. Gen.* 133, 450–467. doi: 10.1037/0096-3445.133.3.450
- Kuo, K., and Vasisht, S. (2006). *Processing Relative Clauses: Evidence from Chinese (Unpublished manuscript)*. Potsdam: University of Potsdam.
- Ledoux, K., Traxler, M. J., and Swaab, T. Y. (2007). Syntactic priming in comprehension: evidence from event-related potentials. *Psychol. Sci.* 18, 135–143. doi: 10.1111/j.1467-9280.2007.01863.x
- Lin, C., and Bever, T. (2006). “Subject preference in the processing of relative clauses in Chinese,” in *Poster Presented at the 25<sup>th</sup> West Coast Conference on Formal Linguistics* (Somerville, MA).
- Liu, T., Zhou, T., and Yang, Y. (2011). Universality of processing advantage for subject relative clause: an ERP study of Chinese relative clauses. *Ling. Sci* 10, 1–20.
- Luka, B. J., and Barsalou, L. W. (2005). Structural facilitation: mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *J. Mem. Lang.* 52, 436–459. doi: 10.1016/j.jml.2005.01.013
- Mehler, J., and Carey, P. (1967). Role of surface and base structure in the perception of sentences. *J. Verbal Learn. Verbal Behav.* 6, 335–338. doi: 10.1016/S0022-5371(67)80122-1
- Mitchell, D., Cuetos, F., Corley, M., and Brybaert, M. (1995). Exposure-based models of human parsing: evidence for the use of coarse-grained (nonlexical) statistical records. *J. Psycholinguist. Res.* 24, 469–488. doi: 10.1007/BF02143162

## FUNDING

This research was supported by Humanities and Social Sciences Youth Foundation of Chinese Ministry of Education (15YJC740091) and National Social Science Foundation of China (13BY069).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00045>

- Pickering, M. J., and Branigan, H. P. (1998). The representation of verbs: evidence from syntactic priming in language production. *J. Mem. Lang.* 39, 633–651. doi: 10.1006/jmla.1998.2592
- Pickering, M. J., Mclean, J. F., and Branigan, H. P. (2013). Persistent structural priming and frequency effects during comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 890–897. doi: 10.1037/a0029181
- Pu, M. (2007). The distribution of relative clauses in Chinese discourse. *Discourse Process.* 43, 25–53. doi: 10.1207/s15326950dp4301\_2
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol. Bull.* 114, 510–532.
- Rayner, K., Carlson, M., and Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *J. Verbal Learn. Verbal Behav.* 22, 358–374. doi: 10.1016/S0022-5371(83)90236-0
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition* 89, 179–205. doi: 10.1016/S0010-0277(03)00119-7
- Scheepers, C., and Crocker, M. W. (2004). “Constituent order priming from listening to comprehension: a visual-world study,” in *The On-Line Study of Sentence Comprehension: Eye Tracking, ERPs, and Beyond*, eds M. Carreiras and C. Clifton. (New York, NY: Psychology Press), 167–185.
- Segaert, K., Menenti, L., Weber, K., and Hagoort, P. (2011). A paradox of syntactic priming: why response tendencies show priming for passives, and response latencies show priming for actives. *PLoS ONE* 6:e24209. doi: 10.1371/journal.pone.0024209
- Spivey-Knowlton, M., and Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227–267. doi: 10.1016/0010-0277(94)00647-4
- Sturt, P., Keller, F., and Dubey, A. (2010). Syntactic priming in comprehension: parallelism effects with and without coordination. *J. Mem. Lang.* 62, 333–351. doi: 10.1016/j.jml.2010.01.001
- Thothathiri, M., and Snedeker, J. (2008a). Give and take: syntactic priming during spoken language comprehension. *Cognition* 108, 51–68. doi: 10.1016/j.cognition.2007.12.012
- Thothathiri, M., and Snedeker, J. (2008b). Syntactic priming during language comprehension in three- and four-year-old children. *J. Mem. Lang.* 58, 188–213. doi: 10.1016/j.jml.2007.06.012
- Tooley, K. M., and Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition* 132, 101–136. doi: 10.1016/j.cognition.2014.04.002
- Tooley, K. M., Traxler, M. J., and Swaab, T. (2009). Electrophysiological and behavioral evidence of syntactic priming in sentence comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 19–45. doi: 10.1037/a0013984
- Traxler, M. J. (2008). Lexically independent priming in online sentence comprehension. *Psychonom. Bull. Rev.* 15, 149–155. doi: 10.3758/PBR.15.1.149
- Traxler, M. J., Pickering, M. J., and Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *J. Mem. Lang.* 39, 558–592. doi: 10.1006/jmla.1998.2600
- Traxler, M. J., and Tooley, K. M. (2008). Priming in sentence comprehension: strategic or syntactic? *Lang. Cogn. Process.* 23, 609–645. doi: 10.1080/01690960701639898
- Trueswell, J. C., Tanenhaus, M. K., and Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang.* 33, 285–318. doi: 10.1006/jmla.1994.1014
- Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 528–553. doi: 10.1037/0278-7393.19.3.528
- Van Gompel, R., and Pickering, M. (2007). “Syntactic parsing,” in *The Oxford Handbook of Psycholinguistics*, ed G. Gaskell (Oxford: Oxford University Press), 289–307.
- Wang, H., and Bing, W. (2013). The incongruity between frequency and processing difficulty of Chinese relative clauses: evidence from ERPs. *Foreign Lang. Res.* 139, 13–22. doi: 10.13978/j.cnki.wyyj.2013.03.007
- Wei, H. (2013). *The Interactive Serial Processing Model: Evidence from the Processing of the Chinese Ambiguous Construction V NP1 de NP2*. Beijing: Science Press.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., and Macdonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cogn. Psychol.* 58, 250–271. doi: 10.1016/j.cogpsych.2008.08.002
- Wu, F. (2009). *Factors Affecting Relative Clause Processing in Mandarin*. (Unpublished doctoral dissertation). University of Southern California, Los Angeles, CA.
- Wu, F., Kaiser, E., and Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Lang. Cogn. Process.* 27, 1489–1524. doi: 10.1080/01690965.2011.614423
- Zagar, D., Pynte, J., and Rativeau, S. (1997). Evidence for early closure attachment on first pass reading times in French. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 50, 421–438.
- Zhang, Q., and Yang, Y. (2010). Object preference in the processing of relative clause in Chinese. *Ling. Sci.* 9, 337–353.
- Zhang, Y., Zhang, H., and Shu, H. (2000). A study on the processing of ambiguous phrases in Chinese. *Acta Psychol. Sinica* 32, 13–19.
- Zhou, T., Zheng, W., Shu, H., and Yang, Y. (2010). The superiority of processing Chinese object-extracted relative clause: evidence from aphasic studies. *Ling. Sci.* 9, 225–243.

**Conflict of Interest Statement:** 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 © 2016 Wei, Dong, Boland and Yuan. 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) or licensor 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.



# A Neurophysiological Investigation of Non-native Phoneme Perception by Dutch and German Listeners

Heidrun Bien<sup>1,2\*</sup>, Adriana Hanulíková<sup>3†</sup>, Andrea Weber<sup>4\*</sup> and Pienie Zwitserlood<sup>2,5</sup>

<sup>1</sup> Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK, <sup>2</sup> Institute for Psychology, University of Münster, Münster, Germany, <sup>3</sup> Albert-Ludwigs-Universität Freiburg, Freiburg, Germany, <sup>4</sup> Eberhard-Karls-Universität Tübingen, Tübingen, Germany, <sup>5</sup> Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, Münster, Germany

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Yang Zhang,  
University of Minnesota, USA  
Stephanie Kathleen Ries,  
University of California, Berkeley, USA

### \*Correspondence:

Heidrun Bien  
h.bien@qmul.ac.uk

†These authors have shared first authorship.

### Specialty section:

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

**Received:** 13 October 2015

**Accepted:** 11 January 2016

**Published:** 29 January 2016

### Citation:

Bien H, Hanulíková A, Weber A  
and Zwitserlood P (2016)  
A Neurophysiological Investigation  
of Non-native Phoneme Perception  
by Dutch and German Listeners.  
*Front. Psychol.* 7:56.  
doi: 10.3389/fpsyg.2016.00056

The Mismatch Negativity (MMN) response has often been used to measure memory traces for phonological representations and to show effects of long-term native language (L1) experience on neural organization. We know little about whether phonological representations of non-native (L2) phonemes are modulated by experience with distinct non-native accents. We used MMN to examine effects of experience with L2-accented speech on auditory brain responses. Specifically, we tested whether it is long-term experience with language-specific L2 pronunciations or instead acoustic similarity between L2 speech sounds that modulates non-native phoneme perception. We registered MMN responses of Dutch and German proficient L2 speakers of English to the English interdental fricative /θ/ and compared it to its non-native pronunciations /s/ (typical pronunciation of /θ/ for German speakers) and /t/ (typical pronunciation of /θ/ for Dutch speakers). Dutch and German listeners heard the English pseudoword *thond* and its pronunciation deviants *sond* and *tond*. We computed the identity Mismatch Negativity (iMMN) by analyzing the difference in ERPs when the deviants were the frequent vs. the infrequent stimulus for the respective group of L2 listeners. For both groups, *tond* and *sond* elicited mismatch effects of comparable size. Overall, the results suggest that experience with deviant pronunciations of L2 speech sounds in foreign-accented speech does not alter auditory memory traces. Instead, non-native phoneme perception seems to be modulated by acoustic similarity between speech sounds rather than by experience with typical L2 pronunciation patterns.

**Keywords:** L2 substitutions, interdental fricative, Dutch, German, non-native phoneme perception, MMN, ERP

## INTRODUCTION

Listeners need to correctly discriminate and identify speech sounds in order to succeed in word recognition. There is ample evidence that experience with a given language influences how listeners perceive, discriminate, and categorize speech sounds (Strange, 1995; Cutler, 2012). This can, for example, be seen when looking at discrimination abilities for phoneme contrasts in the listener's native language (L1) compared to discrimination abilities of unknown contrasts in a second language (L2; e.g., Werker and Tees, 1984). While discrimination in one's native language is usually easy, discrimination success in a L2 is modulated by how well the non-native sounds fit existing



native categories. Indeed, cross-linguistic studies show that different language backgrounds effect L2 speech perception (e.g., Flege, 1995, 2007; Strange, 1995; Best and Tyler, 2007). Models of phonetic perception in L2, such as Flege's Speech Learning Model (SLM; Flege, 1995) and Best's Perceptual Assimilation Model (PAM; Best and Tyler, 2007), therefore predict discriminability of phoneme categories by L2 listeners by reference to the relationship of the phoneme repertoires of their first and second language. While PAM deals with inexperienced listeners, Flege's SLM focuses on experienced L2 learners and predicts increasing difficulties in establishing a new category with a decreasing acoustic-phonetic distance between an L1 and an L2 sound. While neither of these accounts deals with experiential effects from listening to L2-accented speech, they both assign an important role to the phonetic similarity between native and non-native sounds.

Experience also shapes the time course of lexical processing in one's native language. Listeners recognize words that occur frequently in their L1 more easily than infrequent words (Marslen-Wilson, 1987), they also recognize native pronunciation variants, as in English *corp'rate* for *corporate*, faster when these variant forms are frequent than when they are infrequent (e.g., Ranbom and Connine, 2007; Connine et al., 2008). Such processing advantages for frequent variants are often seen as an indicator for what form might be stored and represented in the mental lexicon (e.g., Ranbom and Connine, 2007). Evidence for experiential effects also comes from cross-linguistic studies examining native and non-native listeners' processing of frequent L2 pronunciation variants (Hanulíková and Weber, 2012). In their eye-tracking study, English listeners as well as Dutch and German learners of English differed in the recognition speed of English words in which the initial phoneme /θ/ was substituted by /s/, /f/, or /t/ (e.g., *theft* pronounced as /tɛft/, *theme* as /fi:m/, and *thrill* as /sril/). In a production experiment, Hanulíková and Weber (2010) showed that while /t/, /f/, and /s/ are the three most common /θ/-substitutions, the relative frequency with which they occur differs across the Dutch and German speakers' non-native productions. The dominant /θ/-substitute for German speakers is /s/, while for Dutch speakers it is /t/. Eye-tracking data from Hanulíková and Weber (2012) revealed that recognition ease of non-native variants reflects these distinct production patterns. For example, listeners heard *theft* pronounced as the variant /tɛft/ and saw four printed words on a computer screen: the intended English word (e.g., *theft*), a phonological rhyme competitor (e.g., *left*), and two unrelated distracters (e.g., *kiss* and *mask*). Looking preferences for target words (e.g., the printed word *theft*) matched the language-specific preferences for producing these variants. Dutch listeners fixated the target words most often when hearing variants with the /t/-substitutions, and German listeners did so when hearing the /s/-substitutions. The authors concluded that linguistic experience with L2 pronunciations facilitates recognition of these variant forms in L2 listening. As robust as these effects are, it remains unclear whether they originate from a phonemic or lexical level.

While experiential factors in L1 perception have been well studied, very little is known about the consequences of L2

experience for neural representations of L2 phonemes. Does experience with typical pronunciations of L2 speech sounds lead to cross-linguistically distinct memory traces for non-native phonemes? The goal of the present study is thus to investigate the effects of long-term L2 experience on the nature of phoneme categories of a second language. To this end, we measured auditory brain responses by using the Mismatch Negativity (MMN). Specifically, does language-specific experience, due to the frequency of pronunciation variants of the English voiceless interdental fricative /θ/, result in distinct MMN responses to /θ/-substitutions as a function of specific differences between Dutch and German accented speech?

## Production and Perception of English Interdental Fricatives

The English fricative /θ/ presents great difficulties in production for many learners of English, and even highly proficient L2 learners regularly substitute English /θ/ with other sounds, most often /s/, /t/, and /f/ (for an overview, see Brannen, 2002). The preferences for substitutes depend on the L1 background of L2 speakers (e.g., Brannen, 2002; Hanulíková and Weber, 2010). Hanulíková and Weber (2010) have shown that German learners of English commonly substitute /θ/ with /s/ (29%) and to a much lesser extent with /t/ (7%) or /f/ (5%), while Dutch learners prefer to use /t/ (23%) and to a much lesser extent /s/ (5%) or /f/ (3%); (Note that all three substitutes are phonemes of both Dutch and German). As a consequence, it is reasonable to assume that German learners experience /s/-substitutes (as in /sɛft/ for *theft*) the most, while Dutch speakers are most often presented with /t/-substitutes (as in /tɛft/ for *theft*). In the present study, we therefore focus on the perception of these two most frequent substitutes.

/θ/ and /s/ are acoustically slightly more similar than /θ/ and /t/. From an articulatory viewpoint, /θ/ and /s/ are fricatives, realized with a constriction in the oral tract that causes turbulent airflow. /t/ on the other hand is an oral stop consonant, for which the vocal tract is first blocked, stopping all airflow, before it is released with a burst. /θ/ is characterized by a relatively flat spectrum with no clearly dominating peaks, while alveolar /s/ displays an intense primary spectral peak at higher frequencies (e.g., Jongman et al., 2000). The spectrum of /t/ has a diffuse spread of energy, with peak amplitudes being larger in the high frequencies (e.g., Stevens and Blumstein, 1978). The two groups of Dutch and German L2 learners of English are particularly interesting, because they not only differ in their predominant [θ]-substitutions, but also in the acoustic properties of both /s/ and /t/ in their respective L1 and from their L2 English. Dutch /s/ is less articulatorily tense and has graver friction than German or English (Mees and Collins, 1982; Rietveld and van Heuven, 2001; Hanulíková and Weber, 2010), and /t/ in initial position is aspirated in German (and in English) but unaspirated in Dutch (Lisker and Abramson, 1964; Keating, 1984).

These acoustic similarities and differences between /θ/-/s/ and /θ/-/t/ do not necessarily affect the ability to perceptually discriminate these pairs. Offline discrimination and identification tasks show that non-native listeners can distinguish between /θ/

and /s/ and /θ/ and /t/ quite well (e.g., Hancin-Bhatt, 1994; Cutler et al., 2004; Hanulíková and Weber, 2012). For example, Cutler et al. (2004) have found that Dutch L2 listeners confuse English /θ/ (in 0-db SNR) with /t/ 6.3% and with /s/ 0.4%. Hancin-Bhatt (1994) showed that German listeners in good listening conditions misidentify /θ/ as /t/ 0% and as /s/ 5%. In line with this pattern, Hanulíková and Weber (2012) showed in an AXB task that performance for both /θ/-/s/ and /θ/-/t/ contrasts was high and comparable across Dutch and German listeners (on average 89% correct for the /θ/-/s/ contrast and 90% correct for the /θ/-/t/ contrast). Although Dutch and German listeners can perceptually distinguish between /θ/-/s/ and /θ/-/t/ quite well, their productions show clear preferences toward one of the variants. While these production preferences affect lexical processing, it is less clear whether non-native phoneme perception is affected as well. This raises the question of the level at which such experiential effects arise during processing. Does the frequency of production variants in L2 speech already affect pre-attentive processing of speech sounds at a pre-lexical level? In other words, is the memory representation of /θ/ closer to /s/ for German listeners, and to /t/ for Dutch listeners?

## MMN Studies on Effects of Experience in Speech Perception

An excellent tool to investigate experience-based auditory memory traces is Mismatch Negativity, an early event-related brain potential (ERP) generated in the auditory cortex. It is a negative ERP component that occurs between 150 and 350 ms after the detection of a deviant feature in the stimulus. In a typical MMN design, the so-called standard stimulus (sound, syllable, or word) is presented 80–90% of the time while the so-called deviant stimulus (sound, syllable, or word) is presented 10–20% of the time. It is assumed that the MMN is evoked through a mismatch of the properties of a deviant stimulus and the neural traces in sensory memory consigned by the repeated presentation of a standard stimulus, irrespective of the direction of the subject's attention or task. As such, an MMN design allows the examination of amplitude differences upon the detection of a change between standard and deviant pronunciations.

Since its discovery in the 1970s (Näätänen et al., 1978), MMN has been linked to various aspects of deviant acoustic properties (for an overview, see Shtyrov and Pulvermüller, 2007) such as pitch (e.g., Näätänen and Gaillard, 1983; Jacobsen et al., 2003), stimulus duration (Paavilainen et al., 1991), and loudness (e.g., Keidel and Spreng, 1965). It has been shown that better discrimination of a native or a non-native phonetic contrast is reflected by higher MMN amplitudes (e.g., Winkler et al., 1999; Shafer et al., 2004). Näätänen et al. (1997) were among the first to observe such language-specific phoneme representations using MMN. In their study, Finnish and Estonian participants were presented with the vowel /e/ (used as standard in the MMN design), that is present in both languages, as well as with vowels /ö/, /o/, /õ/ (used as deviants in the MMN design), of which the first two are present in both languages but the last one only exists in Estonian. Näätänen et al. (1997) found that the amplitude of the MMN was influenced by the deviant's phonemic status

in the respective language. There was larger MMN for vowels that were present in the participant's native language (Finnish) compared to vowels that were not present. The effect did not seem to be affected by acoustic features, since the deviant vowels were equally complex. Larger MMN occurred only when the deviant stimulus was part of the respective phoneme inventory. This result led to the suggestion that memory traces of speech sounds are language-dependent, and reflect native phoneme categories (cf. Bien and Zwitserlood, 2013). A replication of the result came from data from 12-months-old but not from 6-months-old infants, suggesting an early development of language-specific memory traces (Cheour et al., 1998).

In the same line of research, Dehaene-Lambertz (1997) found that native French-speaking subjects display MMN when confronted with an acoustic change signaling a phonemic boundary in French but not in Hindi. Effects of experience with a L1 are also visible when experience is operationalized as the relative frequency of occurrence for a given phonological process in a given context. To examine sensitivity to frequency of phonological variants, Tavabi et al. (2009) created German bisyllables and manipulated the phonemic context in which assimilation of /n/ to /m/ occurs (e.g., *onbo* to *ombo*) as well as the frequency of assimilation (/n/ to /m/ is more frequent than /m/ to /n/). They found that both the frequency of the particular assimilation and the context in which it occurs modulate the MMN.

MMN can be used to index the perception and discrimination abilities of foreign-language phonemes as well; Winkler et al. (1999) demonstrated that Hungarian participants with no prior exposure to Finnish showed no MMN and very poor discrimination performance with the Finnish vowel contrast (/e/ – /æ/). Hungarians who were fluent in Finnish showed a MMN that was comparable to the one found in native Finnish-speaking participants. Training effects for the perception of non-native contrasts – Germans learning moraic consonant duration in Japanese – are reflected in the emergence of an MMN (Menning et al., 2002). Interestingly, presenting a continuum of synthesized Hindi stops to English and Hindi speakers, Shafer et al. (2004) found some evidence that pre-attentive discrimination is modulated by experience with the speech sounds of a language. The observed MMN did not, however, directly correspond to the behavioral discrimination results, and some pairs of sounds that could be behaviorally discriminated did not elicit MMN. Long-term experiential factors with L2 phoneme duration were tested by Nenonen et al. (2003). They found that – despite extensive experience with the L2 and advanced L2 skills – non-native listeners did not reach native-like discrimination abilities for speech stimuli (but they were comparable with natives when tested with non-speech stimuli).

Taken together, these studies suggest that the MMN can be used as an index of long-term experience with native and non-native speech sounds, using single speech sounds, syllables, and non-words. In this study, we examine electrophysiological activity of the brain to understand the perception of English dental fricative sounds in two groups of proficient L2 listeners, which has been rarely done. Some previous research (mainly using magnetoencephalography) on fricative perception

examined L1 English phonemic contrasts such as /s/ and /ʃ/ (Miller and Zhang, 2014; Lago et al., 2015) as well as responses to Polish fricatives by native and inexperienced non-native listeners (Lipski and Mathiak, 2007). It remains unclear whether experience with typical mispronunciations of L2 speech sounds lead to cross-linguistically distinct memory traces for non-native phonemes. In our study, we examined this question by using English monosyllables with no lexical status to avoid possible top-down effects (Pulvermüller and Shtyrov, 2006), and to focus on L2 memory traces for phonemes.

## Present Study

In the present study, we use MMN to look at the role of experience with common mispronunciations in a second language. Specifically, we examine whether cross-linguistically distinct experience with mispronunciations of L2 speech sounds shapes the neural organization of L2 phonemes, as reflected in the size of mismatch effects. Studying Dutch and German participants, we focus on the perception of the voiceless interdental fricative /θ/ and its substitutions /t/ and /s/, most commonly produced by these two groups of learners of English.

To examine the influence of experience with non-native accents on auditory memory traces, we compared the automatic electrophysiological responses in Dutch and German listeners to the English pseudoword *thond* and its pronunciation variants *sond* and *tond* in an oddball paradigm. Oddball paradigms are not free from lexical effects, even with attention diverted from the acoustic stimuli (cf. Pulvermüller and Shtyrov, 2006). Therefore, we used English monosyllabic pseudowords. We concentrated on the variant forms /s/ and /t/ as they represent the preferred substitute for the two learner groups respectively. If long-term experience with typical non-native variants already affects this early automatic processing level, we should find a similar accent-specific pattern of results as reported for lexical processing in Hanulíková and Weber (2012). That is, smaller mismatch effects should be found for *tond* than for *sond* for Dutch listeners, and the reverse should be found for German listeners. Alternatively, exposure-frequency effects might only arise at higher levels of lexical processing and might not affect non-native phoneme representations. In that case, the two variant forms *tond* and *sond* might either elicit comparable brain responses, or they might reflect effects of stimulus similarity, in which case *tond* should elicit a larger mismatch effect than *sond* for both Dutch and German participants.

## MATERIALS AND METHODS

### Participants

Eighteen native speakers of Dutch (mean age: 23, *SD*: 3.3, nine male) and 17 native speakers of German (mean age: 23, *SD*: 1.6, three male) participated in the present study, after having given written, informed consent. Dutch participants were tested in the Netherlands, at the Max Planck Institute for Psycholinguistics. German participants were tested in Germany, at the University of Münster. All participants reported having normal hearing and no history of neurological problems, head injuries, or continuous

medication. Participation was compensated with €12 or course credit.

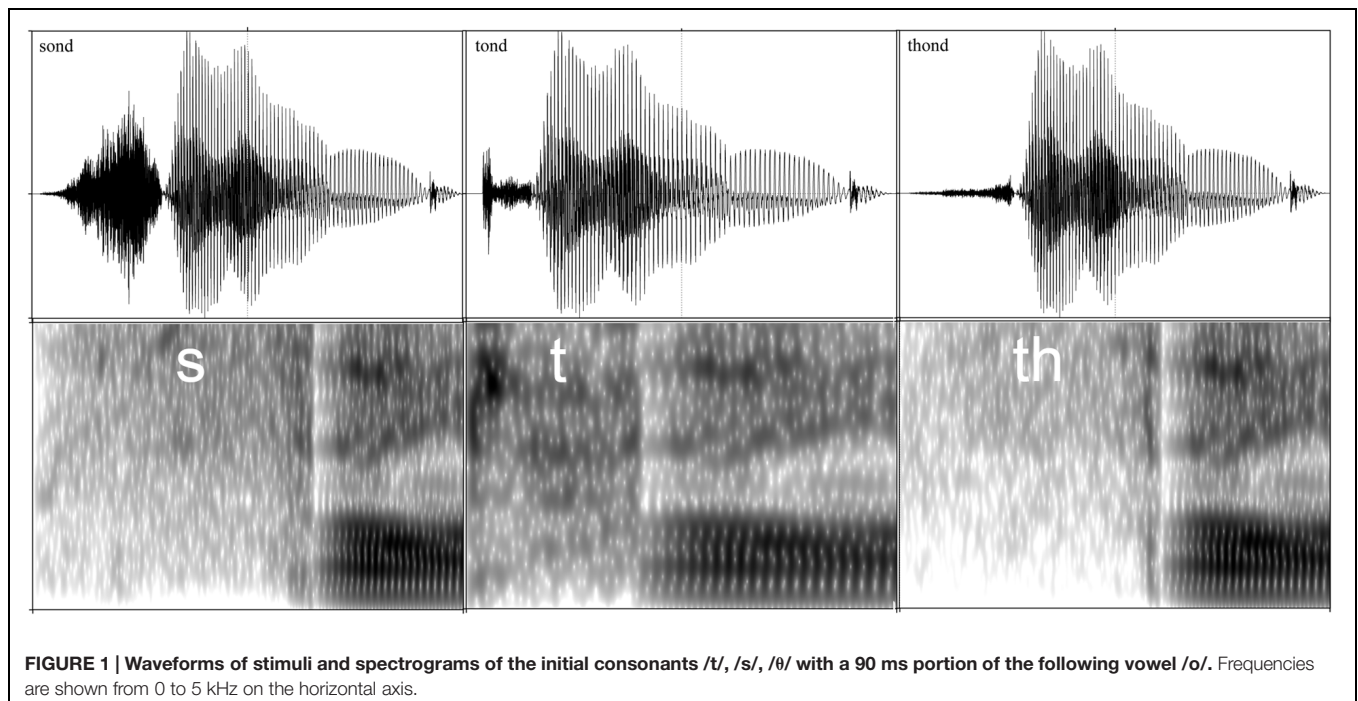
Subsequent to the experiment, participants took part in an ABX discrimination test of the speech materials, and provided information on their use of and proficiency in English. All German participants had learned English in school as their second language with a mean duration of 8.4 years (*SD*: 0.8). Dutch participants had on average 7.6 years (*SD*: 0.7) of English education in school. In the Netherlands, all students in upper educational levels have to attend German language courses for at least 3 years, and German is usually their third or fourth non-native language (after English). Thus, all Dutch participants had some knowledge of German. Dutch, on the other hand, is not mandatory in German high schools, and German participants had little or no exposure to Dutch.

This study was carried out in accordance with the recommendations for ethical guidelines of the Institute for Psychology, Westfälische Wilhelms-Universität, Münster, Germany and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. All participants gave written informed consent in accordance with the Declaration of Helsinki.

### Stimuli and Design

We compared processing of the English interdental fricative /θ/ in the pseudoword *thond* with the non-native pronunciation variants *tond* and *sond*. The variant *sond* represents a typical pronunciation of *thond* for German speakers of English, who frequently substitute /θ/ with /s/, while the variant *tond* is typical for Dutch speakers of English, who frequently substitute /θ/ with /t/ (cf. Hanulíková, and Weber, 2010). The stimuli used in the experiment were therefore the English monosyllabic pseudowords *thond*, *sond*, and *tond*. To ensure a native-like pronunciation of /θ/ in *thond*, all pseudowords were produced by a native speaker of English. None of the stimuli is, or closely resembles, an existing word in Dutch or in German. In addition, pronounced as English pseudowords, *thond*, *sond*, and *tond* cannot be interpreted as Dutch or German pseudowords, due to a violation of the phonotactic constraint of syllable-final devoicing (e.g., in Dutch and in German, the pseudoword *sond* would be pronounced /sont/). The length of the initial consonants was 149 ms for *thond*, 60 ms for *tond* and 176 ms for *sond*. The length of the stimuli was 593 ms for *thond*, 499 ms for *tond*, and 609 ms for *sond*. The stimuli were cross- and identity-spliced to avoid eliciting MMN due to features other than the initial phoneme in the recorded materials (see **Figure 1** for stimuli waveforms and spectrograms after the splicing procedure). Some variation in the stimuli was re-created by changing the pitch to abstract away from specific acoustic properties of individual tokens (e.g., Bien et al., 2009). With three stimuli and five levels of pitch (+12, +6, +0, -6, and -12 Hz), the total number of tokens was 15. All stimuli served as both standards and deviants in different blocks.

The experiment consisted of four blocks, each with a different STANDARD\_deviant combination ([THOND\_tond]; [THOND\_sond]; [TOND\_thond]; [SOND\_thond]). The order of blocks was balanced across participants. Within each block, 500 stimuli were presented in random order, with a deviant



likelihood of 20% and an inter-stimulus interval of 1000 ms. Each block lasted for approximately 11 min, and there was a short break after each block. The experimenter started the next block once the participants had retaken a stable and comfortable position.

### Procedure and EEG Recording

Participants were comfortably seated in front of a computer screen in a sound-attenuated room. The stimuli were presented via loud speakers at approximately 60 dB SPL. During the electrophysiological recordings, participants watched a silent movie and were told that they could ignore the auditory stimuli.

The electroencephalography (EEG) of the German participants (GER) were recorded in sampling rates of 256 Hz, using 64-channel WaveGuard caps (ANT, Enschede, Netherlands) connected to an ANT amplifier (ANT, Enschede, NL). AFz was used as the ground electrode, and electrode impedances were kept below 5 KΩ. Horizontal eye movements were recorded using two bipolar electrodes with left and right canthal montage. Lateral eye movements and blinks were recorded using two bipolar electrodes placed above and below the right eye. An average mastoid reference was used.

Electroencephalography for the Dutch participants (NL) was recorded from 34 Ag–AgCl electrodes (Brain Products, MedCat, Netherlands) at standard 10–20 locations. Impedance was kept below 5 kΩ. All recordings were referenced to the left mastoid during recording (eye movement and blink artifacts were recorded from F9 to F10 and from Fp1 to an additional EOG electrode below the left eye), amplified with BrainAmp DC amplifiers (0.016–100 Hz band pass, digitized at 500 Hz), and re-referenced off-line to the mastoid average (e.g., Poellmann, 2013; Poellmann et al., under revision).

Data were analyzed with Advanced Source Analyses (ASA) software (ANT Software BV, Enschede, NL) and with SPSS statistics. We filtered the data offline, applying a 35 Hz low-pass filter. EEGs outside the range of  $-75$  to  $+75$   $\mu$ V were labeled as artifacts and excluded from further analyses. This ensured the elimination of segments containing eye movement, blinking, or muscular activity. Overall, 81% of epochs were free from artifacts and used for analyses (71% for German and 91% for Dutch participants). Intact epochs were evenly distributed across conditions within each group. The remaining data were averaged in epochs of 800 ms, including a 250 ms pre-stimulus baseline interval used for epoch correction. All analyses were based on the mean amplitudes at Cz.

We analyzed the *identity* mismatch (*i*MMN) elicited by the *thond*-pronunciation deviants *tond* and *sond* in the Dutch and German participants. In order to compute the *i*MMN for *tond* and *sond*, we subtracted the respective ERPs when used as a standard next to *thond* from the ERPs when used as a deviant next to *thond*. That is, for the *i*MMN of *tond*, the standard-ERP elicited by *tond* in the block [TOND\_Thond] was subtracted from its deviant-ERP elicited in the block [THOND\_tond]. For *sond*, the standard-ERP in block [SOND\_Thond] was subtracted from its deviant-ERP elicited in block [THOND\_sond]. This *i*MMN procedure is specifically relevant when stimuli differ with respect to duration and various spectral factors, which certainly holds for fricatives and plosives. Calculating *i*MMN cancels out the specifics of the individual acoustic stimulus tokens (**Figure 2**).

Based on qualitative visual inspection, the time windows for the analyses of the mismatch effects were determined by the range of the deviant-N1. Note that mismatch negativity often overlaps with the N1 (cf. Schröger, 1998). The data-driven selection of the time windows was done separately for

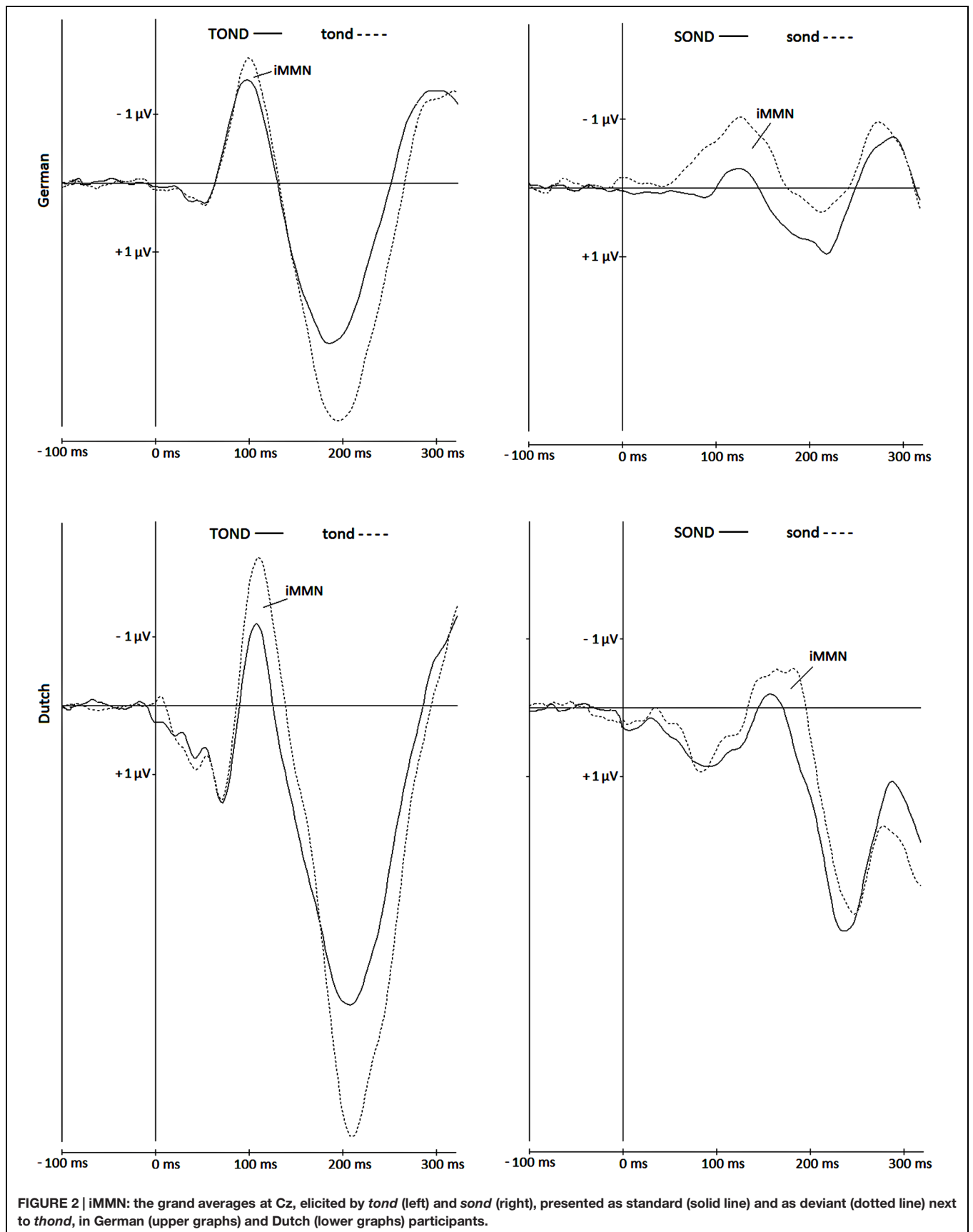


FIGURE 2 | iMMN: the grand averages at Cz, elicited by *tond* (left) and *sond* (right), presented as standard (solid line) and as deviant (dotted line) next to *thond*, in German (upper graphs) and Dutch (lower graphs) participants.

the Dutch and German groups of listeners. Measured at Cz (Figure 2), where both the N1s and the mismatch effects were most prominent, the time windows were 88–140 ms (Dutch) and 68–133 ms (German) for *tond*, and 134–196 ms (Dutch) and 78–168 msec (German) for *sond*. The use of different time windows is justified for the factor *phoneme* because of the large variance in the onset of perceivable information between the critical stimuli. Likewise, because listeners' perception is optimized for their native language, Dutch and German listeners differ with respect to the uptake of information that distinguishes between phonemes (the voice-onset times vs. prevoicing distinction for voiced plosives is a prime example). Thus, to select most objectively for the planned comparison of the identity mismatch elicited by a given stimulus, we also opted for data-driven (and thus potentially different) time windows for the factor *listener group*. Note that it is not uncommon to use data-driven solutions to the problem of latency variability (see Luck, 2005, p. 135). All analyses were based on the mean amplitudes at Cz within the specified time windows (i.e., over the whole range of the deviant-N1). We followed the suggestions by Luck (2005) to use an area amplitude measure rather than a peak amplitude measure to mitigate the reduction in amplitude caused by latency variability amplitude (see also Schröger, 1998). For the statistical analysis, we used a 2 (Deviance: *sond*, *tond*) by 2 (Group: Dutch, German) repeated-measures analyses of variance (ANOVA) and report Greenhouse–Geisser corrections and corrected F-values where appropriate. Additionally, we report one-way ANOVAs for Deviance for each group separately.

### Behavioral ABX-Experiment

After the EEG experiment, all participants completed a speech-sound discrimination ABX-task to test participants' discriminatory ability for the three stimuli. Stimuli /θond/, /tond/, and /sond/ were presented over speakers in 12 trials in a random order at the A and B positions, followed by a third stimulus at the X position that matched either A or B. Participants had to press the left shift button when the last presented (X) stimulus matched the first (A) stimulus, and to press the right shift button when it matched the second (B) stimulus. Stimuli were presented at ISIs of 800 ms.

The results showed that German participants distinguished /θond/ equally well from /tond/ (2,9% errors) and /sond/ (4,4% errors). For Dutch participants, it was harder to distinguish /θond/ from /sond/ (22,2% errors) than from /tond/ (8,3% errors). A closer look at the Dutch participants shows that the higher error rate is mainly due to three participants. The error rate drops to 9,3% when these participants are excluded. Note that this discrimination pattern would go against the predicted experience-based perception effect in the EEG study, according to which Dutch speakers would perceive /t/ as a closer match to /θ/ than to /s/ (as reflected in their production behavior). Moreover, previous studies reported non-significant MMN responses to L2 contrasts in L1 and L2 participants despite the presence of differences in behavioral discriminations of L2 contrasts (e.g., for the Japanese listeners' difficulties with the English /r/ and /l/; Zevin et al., 2010).

## RESULTS

Visual inspection of the data in Figure 2 indicated that, when presented next to *thond*, both pronunciation variants *sond* and *tond* elicited mismatch effects in both groups of participants. However, against our hypothesis, the numerical differences of mismatch effects for /s/ and /t/ between the language groups show larger effects for /tond/ in Dutch listeners, and larger effects for /sond/ in German listeners.

In a first step, we tested the significance of each identity mismatch component against zero in each group of participants. Subtracting the standard-ERP of *sond*, elicited in block [SOND\_thond], from its deviant-ERP, elicited in block [THOND\_sond], the identity mismatch effect (mean amplitude in the specified time window) was  $-0.65 \mu\text{V}$  [ $t(1,17) = 4.51, p = 0.049$ ] for Dutch participants and  $-0.81 \mu\text{V}$  [ $t(1,16) = 7.08, p = 0.017$ ] for German participants. The difference in the *sond*-iMMNs between the two groups was non-significant (independent sample *t*-test [ $t(33) = 0.379, p = 0.707$ ]). *Tond* elicited iMMNs of  $-0.93 \mu\text{V}$  [ $t(1,17) = 8.31, p = 0.010$ ] in Dutch participants and  $-0.20 \mu\text{V}$  [ $t(1,16) = 0.44, p = 0.519$ ] in German participants. The difference in the *tond*-iMMNs did not reach significance (independent-sample *t*-test [ $t(33) = -1.667, p = 0.105$ ]).

In a second step, ANOVAs on the identity mismatch effects were carried out separately for the two groups of listeners; the factor Deviance (*sond*, *tond*) was not significant in either the Dutch [ $F(1,17) = 0.35, p = 0.563$ ] or German group [ $F(1,16) = 1.88, p = 0.189$ ]. An overall ANOVA with Group (Dutch, German) as the between-subjects factor and Deviance as the within-subjects factor revealed no main effect of Deviance [ $F(1,33) = 0.027, p = 0.609$ ], no main effect of group [ $F(1,33) = 0.961, p = 0.334$ ], and no significant interaction between Group and Deviance [ $F(1,33) = 1.88, p = 0.180$ ; see Table 1].

To summarize, mismatch effects were seen for both deviant stimuli in both language groups (except for the iMMN for *tond* in the German group, where it was expected to be pronounced). Interestingly, the pattern of identity mismatch elicited by the *thond*-pronunciation variants *sond* and *tond* was comparable in both groups of listeners (see Table 1). This does not confirm

**TABLE 1 | Analyses of variance on the identity mismatch effects (mean amplitudes in  $\mu\text{V}$ ) elicited by the *thond*-pronunciation variants *tond* and *sond*, computed with GROUP (Dutch, German) as a between-subject factor, and for each group separately.**

Dutch listeners		
DEVIANCE ( <i>sond</i> , <i>tond</i> )	$F(1,17) = 0.35$	$p = 0.563$
<i>tond</i> $-0.93 \mu\text{V}$	$t(17) = 8.31$	$p = 0.010$
<i>sond</i> $-0.65 \mu\text{V}$	$t(17) = 4.51$	$p = 0.049$
German listeners		
DEVIANCE ( <i>sond</i> , <i>tond</i> )	$F(1,16) = 1.88$	$p = 0.189$
<i>tond</i> $-0.20 \mu\text{V}$	$t(16) = 0.44$	$p = 0.519$
<i>sond</i> $-0.81 \mu\text{V}$	$t(16) = 7.08$	$p = 0.017$
Overall		
GROUP (Dutch, German)	$F(1,33) = 0.96$	$p = 0.334$
DEVIANCE ( <i>sond</i> , <i>tond</i> )	$F(1,33) = 0.27$	$p = 0.609$
DEVIANCE * GROUP	$F(1,33) = 1.88$	$p = 0.180$

an L2-accent-specific pattern of results, according to which smaller mismatch effects were expected for *tond* than for *sond* for Dutch listeners, and the reverse was expected for German listeners. What we observed instead is that the two variant forms *tond* and *sond* elicited comparable brain responses across the two listener groups and thus might reflect effects of stimulus similarity.

## DISCUSSION

The present study examined whether pre-attentive processing of pronunciation variants in non-native speech is influenced by cross-linguistically distinct experiences with such variants. If experience exerts a predominant influence on speech processing and speech-sound representation, smaller mismatch effects were expected for *tond* compared to *sond* in Dutch listeners, for whom /t/ is the common substitute for /θ/ (Hanulíková and Weber, 2010). The reverse was expected for Germans, who frequently substitute /θ/ with /s/ (Hanulíková and Weber, 2010). While there is converging evidence that experience with pronunciation variants in an L2 influences speech processing at a lexical level (e.g., Hanulíková and Weber, 2012), the present study found no evidence for an impact of experience with L2 pronunciations on L2 phoneme representations. We did not find (at least in the ANOVA analysis) the predicted differential processing between the two groups. Note that the numerical differences of mismatch effects for /s/ and /t/ between the language groups were even against the hypothesized direction, with larger effects for /tond/ in Dutch listeners, and for /sond/ in German listeners. A possible explanation is provided further below, however, given the lack of interaction, these differences should not be interpreted.

In the ERPs of Dutch and German proficient speakers of English, we compared the identity mismatch effects elicited by the pronunciation variants *sond* and *tond* in the context of the English pseudoword *thond*. Presented next to the pseudoword *thond*, the pronunciation variants *sond* and *tond* elicited mismatch effects in both the Dutch and German groups of listeners. This was evident when using the identity mismatch approach (presenting two blocks with switching roles of standard and deviant for each variation of interest, then subsequently comparing the ERPs elicited by the identical stimulus when presented as standard and when presented as deviant).

Due to its clearer acoustic onset, *tond* elicited a more distinct and more negative N1 than *sond*, both when presented as a standard and when presented as a deviant. This acoustic difference, however, was not confirmed in a main effect of Deviance in the iMMN. Most importantly, the mismatch effects elicited by *sond* and *tond* were statistically comparable, and there were no interactions with listener group. This result seems to reflect effects of stimulus similarity comparable across both listener groups. Note that we used English monosyllabic pseudowords (*thond*, *tond*, *sond*) and avoided lexical items because pronunciation substitutions may vary depending on the phonemic context and position in a given word. An interesting question for future research is thus whether the same result would be obtained for real words.

It should be noted, however, that while the mean latencies of the time frames were similar between the two listener groups in the *tond*-condition, they were different in the *sond*-condition. One possible reason for this difference could stem from a different uptake of acoustic cues for the English fricative between the two listener groups. Previous studies have shown that Dutch /s/ is articulatorily less tense and has graver friction than German /s/ (Mees and Collins, 1982) and therefore also differs from English to a larger extent than German (e.g., Hanulíková and Weber, 2010, 2012). Slight latency differences present in the *tond*-condition could also be explained by the distinct L1 acoustic characteristics. The Dutch /t/ is less aspirated than the English /t/ and mainly uses prevoicing as a cue to voicing, while German listeners use the VOT to categorize voicing of plosives (Lisker and Abramson, 1964; Keating, 1984). This could lead to distinct /t/-categorization patterns for German compared to Dutch listeners. These differences could have perceptual consequences and this could explain why the mapping of the English /s/ and /t/ sounds onto the distinct Dutch /s/ and /t/ resulted in different latencies than mapping of the English sound onto the more similar German /s/ and /t/. Distinct uptake processes of acoustic information are very likely despite behavioral tasks suggesting that both Dutch and German listeners show comparably high discriminatory performance for both /θ/-/s/ and /θ/-/t/ contrasts (e.g., Hanulíková and Weber, 2012). Although the control ABX test run after the present EEG study showed a higher error rate for /θ/-/s/ compared to /θ/-/t/ in the group of Dutch participants, this was mainly due to three participants. It would be interesting for future research to examine the issue of how cues are weighted differently in the foreign and native languages, particularly in the two highly related languages such as German and Dutch.

Finally, the time windows for analyzing the auditory components were selected based on the grand average of the deviant-N1s. Though we selected the time windows separately for *tond* and *sond* and separately for the group of Dutch and German participants, selection was not based on the individual responses of each participant. The time spans of certain ERP components can vary greatly across individuals (e.g., Michaelewski et al., 1986). As a consequence, group average analyses can eliminate individual mismatch effects, underestimating their actual size. The observation of significant effects based on group averages can be considered a strong indicator that these effects are real. Moreover, we restricted our analysis to the first N1 elicited by the speech sound, time-locked to its onset. Another approach would be to also analyze the second N1, elicited by the acoustic change complex defined as – and time-locked to – the transition of the consonant to the following vowel (see Lipski and Mathiak, 2007; Miller and Zhang, 2014). The ERP data in both subject groups (see **Figure 2**) indicate the consistent elicitation of the negative deflection responses to fricatives. Future research might want to focus on the time-point-by-time-point ERP responses to the speech stimuli to accommodate the complexity of ERP waveforms that could arise from tracking the acoustic properties of speech stimuli in the time domain.

A number of MMN studies found language-specific effects on phoneme perception (e.g., Dehaene-Lambertz, 1997; Näätänen et al., 1997; Winkler et al., 1999; Jacobsen and Schröger, 2003;

Nenonen et al., 2003). However, most of these studies examined native listeners' perception or L2 perception of familiar and unfamiliar contrasts with within- or across-category manipulation for a given language. The present study used phonemes that are frequently produced by non-native speakers of English as substitutions for interdental fricatives. The present study is thus the first to look at whether cross-linguistically distinct experience with frequent non-native pronunciations modulates phoneme representations. The results suggest that experience-based memory representations for frequent or preferred phoneme substitutions are not (or not well) established. Given previous research on differences between native and advanced learners' perception of acoustic features (e.g., Nenonen et al., 2003), the present result may not be surprising. The formation of stable language-specific memory representations in an L2 seems to require early exposure or, for late learners, frequent exposure and practice. Listeners can learn to distinguish non-native contrasts, even when they are unable to produce these (Menning et al., 2002). It would be good if learning novel L2 speech sounds would result in the formation of a new perceptual category (for /θ/, for example), even if correct production lags behind. The use of L2 pronunciation variants may facilitate lexical processing, but does not lead to an annexation of novel sounds into native categories. Indeed, cross-linguistic perception studies and models (e.g., SLM, PAM) suggest that it is more difficult to establish a new L2 phonological category when the acoustic-phonetic properties of an L2 sound are similar to an L1 sound (Flege, 1995, 2007; Best and Tyler, 2007; see Dobel et al., 2009, for evidence from the N400 component). While these models are not directly concerned with L2-accented speech, the present results could be explained on the basis of

acoustic/auditory properties of the consonant pairs (/θ/ as more similar to /s/ than to /t/).

Taken together, the results suggest that long-term non-native experience with frequent pronunciation variants in a second language does not alter memory traces of phonemes or perception of these L2 speech sounds. Instead, pre-attentive perception of non-native speech sounds may better be explained in terms of acoustic similarity to native categories.

## ACKNOWLEDGMENTS

We wish to thank Merel van Rees Vellinga for her help during the experiment set-up and data collection in the Netherlands, and Katharina Dohm and Sophia Thrun for her assistance with data collection in Germany. We thank both reviewers for their helpful and constructive comments on a previous version of this paper. This research was funded by the Max-Planck Society, Germany.

## FUNDING

The article processing charge was funded by the open access publication fund of the Albert Ludwigs University Freiburg.

## AUTHOR CONTRIBUTIONS

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

## REFERENCES

- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds M. J. Munro and O.-S. Bohn (Philadelphia: John Benjamins), 13–34. doi: 10.1075/llt.17.07bes
- Bien, H., Lagemann, L., Dobel, C., and Zwitserlood, P. (2009). Implicit and explicit categorization of speech sounds – dissociating behavioural and neurophysiological data. *Eur. J. Neurosci.* 30, 339–346. doi: 10.1111/j.1460-9568.2009.06826.x
- Bien, H., and Zwitserlood, P. (2013). Processing nasals with and without consecutive context phonemes: evidence from explicit categorization and the N100. *Front. Psychol.* 4:21. doi: 10.3389/fpsyg.2013.00021
- Brannen, K. (2002). The role of perception in differential substitution. *Can. J. Linguist.* 47, 1–20.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, J., Allik, K., Alho, R., et al. (1998). Development of language-specific phoneme representations revealed by electric and magnetic brain responses. *Nat. Neurosci.* 1, 351–353. doi: 10.1038/1561
- Connine, C. M., Ranbom, L. J., and Patterson, D. J. (2008). Processing variant forms in spoken word recognition: the role of variant frequency. *Percept. Psychophys.* 70, 403–411. doi: 10.3758/PP.70.3.403
- Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. Cambridge, MA: MIT Press.
- 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
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport* 8, 919–924. doi: 10.1097/00001756-199703030-00021
- Dobel, C., Lagemann, L., and Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3697–3709. doi: 10.1098/rstb.2009.0158
- Flege, J. (1995). "Second language speech learning: theory, findings and problems," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, ed. W. Strange (Baltimore: York Press), 233–277.
- Flege, J. (2007). "Language contact in bilingualism: phonetic system interactions," in *Laboratory Phonology*, Vol. 9, eds J. Cole and J. Hualde (Berlin: Mouton de Gruyter), 353–382.
- Hancin-Bhatt, B. (1994). Segmental transfer: a natural consequence of a dynamic system. *Second Lang. Res.* 10, 242–270. doi: 10.1177/026765839401000304
- Hanulíková, A., and Weber, A. (2010). "Production of English interdental fricatives by Dutch, German, and English speakers," in *New Sounds: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech*, eds K. Dziubalska-Kořaczyk, M. Wrembel, and M. Kul (Poznan: Adam Mickiewicz University), 173–178.
- Hanulíková, A., and Weber, A. (2012). Sink positive: linguistic experience with th-substitutions influences nonnative word recognition. *Attent. Percept. Psychophys.* 74, 613–629. doi: 10.3758/s13414-011-0259-7
- Jacobsen, T., and Schröger, E. (2003). Measuring duration mismatch negativity. *Clin. Neurophysiol.* 114, 1133–1143. doi: 10.1016/S1388-2457(03)00043-9
- Jacobsen, T., Schröger, E., Horenkamp, T., and Winkler, I. (2003). Mismatch negativity to pitch change: varied stimulus proportions in controlling effects of neural refractoriness on human auditory event-related brain potentials. *Neurosci. Lett.* 344, 79–82. doi: 10.1016/S0304-3940(03)00408-7



- Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108, 1252–1263. doi: 10.1121/1.1288413
- Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286–319. doi: 10.2307/413642
- Keidel, W. D., and Spreng, M. (1965). Neurophysiological evidence for the Stevens power function in man. *J. Acoust. Soc. Am.* 38, 191–195. doi: 10.1121/1.1909629
- Lago, S., Scharinger, M., Kronrod, Y., and Idsardi, W. J. (2015). Categorical effects in fricative perception are reflected in cortical source information. *Brain Lang.* 143, 52–58. doi: 10.1016/j.bandl.2015.02.003
- Lipski, S. C., and Mathiak, K. (2007). A magnetoencephalographic study on auditory processing of native and nonnative fricative contrasts in Polish and German listeners. *Neurosci. Lett.* 415, 90–95. doi: 10.1016/j.neulet.2007.01.001
- Lisker, L., and Abramson, A. S. (1964). A cross-language study of voicing in initial stops-acoustical measurements. *Word* 20, 384–422. doi: 10.1080/00437956.1964.11659830
- Luck, S. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9
- Mees, I., and Collins, B. (1982). A phonetic description of the consonant system of Standard Dutch (ABN). *J. Int. Phon. Assoc.* 12, 2–12. doi: 10.1017/S0025100300002358
- Menning, H., Imaizumi, S., Zwitserlood, P., and Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learn. Mem.* 9, 253–267. doi: 10.1101/lm.49402
- Michalewski, H. J., Prasher, D. K., and Starr, A. (1986). Latency variability and temporal interrelationships of the auditory event-related potentials (N1, P2, N2 and P3) in normal subjects. *Electroencephalogr. Clin. Neurophysiol.* 65, 59–71. doi: 10.1016/0168-5597(86)90037-7
- Miller, S., and Zhang, Y. (2014). Neural coding of phonemic fricative contrast with and without hearing aid. *Ear Hear.* 35, e122–e133. doi: 10.1097/AUD.0000000000000025
- Näätänen, R., and Gaillard, A. W. K. (1983). “The orienting reflex and the N2 deflection of the event-related potential (ERP),” in *Tutorials in Event Related Potential Research: Endogenous Components*, eds A. W. K. Gaillard and W. Ritter (Amsterdam: North Holland), 119–141. doi: 10.1016/S0166-4115(08)62036-1
- Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–353. doi: 10.1016/0001-6918(78)90006-9
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huutilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0
- Nenonen, S., Shestakova, A., Huutilainen, M., and Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cogn. Brain Res.* 16, 492–495. doi: 10.1016/S0926-6410(03)00055-7
- Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., and Näätänen, R. (1991). Right-hemisphere dominance of different mismatch negativities. *Electroencephalogr. Clin. Neurophysiol.* 78, 466–479. doi: 10.1016/0013-4694(91)90064-B
- Poellmann, K. (2013). *The Many Ways Listeners Adapt to Reductions in Casual Speech*. Ph.D. thesis, Radboud University Nijmegen, Nijmegen.
- Pulvermüller, F., and Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79, 49–71. doi: 10.1016/j.pneurobio.2006.04.004
- Ranbom, L. J., and Connine, C. M. (2007). Lexical representations of phonological variation in spoken word recognition. *J. Mem. Lang.* 57, 273–298. doi: 10.1016/j.jml.2007.04.001
- Rietveld, A. C. M., and van Heuven, V. J. (2001). *Algemene Fonetiek*. Bussum: Coutinho.
- Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behav. Res. Methods Instrum. Comput.* 30, 131–145. doi: 10.3758/BF03209423
- Shafer, V. L., Schwartz, R. G., and Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cogn. Brain Res.* 18, 242–254. doi: 10.1016/j.cogbrainres.2003.10.007
- Shtyrov, Y., and Pulvermüller, F. (2007). Language in the mismatch negativity design. Motivations, benefits, and prospects. *J. Psychophysiol.* 21, 176–187. doi: 10.1027/0269-8803.21.34.176
- Stevens, K. N., and Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 64, 1358–1368. doi: 10.1121/1.382102
- Strange, W. (ed.). (1995). *Speech Perception, and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press.
- Tavabi, K., Elling, L., Dobel, C., Pantev, C., and Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: a magnetoencephalography study. *PLoS ONE* 4:1–7. doi: 10.1371/journal.pone.0004452
- Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3
- Winkler, I., Kujala, T., Titinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642. doi: 10.1111/1469-8986.3650638
- Zevin, J. D., Datta, H., Maurer, U., Rosania, K. A., and McCandliss, B. D. (2010). Native language experience influences the topography of the mismatch negativity to speech. *Front. Hum. Neurosci.* 4:212. doi: 10.3389/fnhum.2010.00212

**Conflict of Interest Statement:** 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 © 2016 Bien, Hanulíková, Weber and Zwitserlood. 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) or licensor 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.



# Processing Preference Toward Object-Extracted Relative Clauses in Mandarin Chinese by L1 and L2 Speakers: An Eye-Tracking Study

Yao-Ting Sung<sup>1,2</sup>, Jung-Yueh Tu<sup>3\*</sup>, Jih-Ho Cha<sup>2</sup> and Ming-Da Wu<sup>2</sup>

<sup>1</sup> Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan, <sup>2</sup> Center of Learning Technology for Chinese, National Taiwan Normal University, Taipei, Taiwan, <sup>3</sup> International Chinese Education Center, School of Humanities, Shanghai Jiao Tong University, Shanghai, China

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Erich David Jarvis,  
Duke University Medical Center, USA  
Julie E. Boland,  
University of Michigan, USA

### \*Correspondence:

Jung-Yueh Tu  
jytu@sjtu.edu.cn

### Specialty section:

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

Received: 15 July 2015

Accepted: 03 January 2016

Published: 21 January 2016

### Citation:

Sung Y-T, Tu J-Y, Cha J-H and  
Wu M-D (2016) Processing  
Preference Toward Object-Extracted  
Relative Clauses in Mandarin Chinese  
by L1 and L2 Speakers: An  
Eye-Tracking Study.  
Front. Psychol. 7:4.  
doi: 10.3389/fpsyg.2016.00004

The current study employed an eye-movement technique with an attempt to explore the reading patterns for the two types of Chinese relative clauses, subject-extracted relative clauses (SRCs) and object-extracted relative clauses (ORCs), by native speakers (L1), and Japanese learners (L2) of Chinese. The data were analyzed in terms of gaze duration, regression path duration, and regression rate on the two critical regions, head noun, and embedded verb. The results indicated that both the L1 and L2 participants spent less time on the head nouns in ORCs than in SRCs. Also, the L2 participants spent less time on the embedded verbs in ORCs than in SRCs and their regression rate for embedded verbs was generally lower in ORCs than in SRC. The findings showed that the participants experienced less processing difficulty in ORCs than SRCs. These results suggest an ORC preference in L1 and L2 speakers of Chinese, which provides evidence in support of linear distance hypothesis and implies that the syntactic nature of Chinese is at play in the RC processing.

**Keywords:** relative clauses, Mandarin Chinese, L2 sentence processing, eye-movements, Japanese CSL learners

## INTRODUCTION

Relative clauses (RCs) have received considerable attention in psycholinguistic and linguistic research over the past few decades. An RC is a subordinate clause that modifies a noun and is embedded within a noun phrase. There are two major types of RCs: subject-extracted relative clauses (SRCs) and object-extracted relative clauses (ORCs). Examples of an SRC and an ORC in English are given in (1a) and (1b), respectively.

- (1a) English SRC  
The principal<sub>i</sub> [who  $e_i$  introduced the teacher]<sub>RC</sub> talked in a very polite manner  
FILLER REL GAP  
“The principal who introduced the teacher talked in a very polite manner.”
- (1b) English ORC  
The principal<sub>i</sub> [who the teacher introduced  $e_i$ ]<sub>RC</sub> talked in a very polite manner  
FILLER REL GAP  
“The teacher whom the principal introduced talked in a very polite manner.”

In (1), “the principal” is extracted from the clause and leaves an empty position, which is called a *gap*. The relative pronoun “who” introduces the RC. The extracted noun phrase “the principal” is coindexed with the gap and is called the *filler*, because it should fill the gap. The two types of RCs only contrast each other with respect to the location of the gap. Hence, comprehending and integrating RCs requires dependency between the filler and gap to be developed in harmony.

SRCs are considered easier to process than ORCs, with the evidence coming from observations of RC processing by native speakers (L1) of head-initial languages (e.g., for English, see Gordon et al., 2006; for Dutch, see Frazier, 1987; for French, see Holmes and O’Regan, 1981; for German, see Schriefers et al., 1995). The SRC preference is also evident in head-final languages (e.g., for Japanese, see Ueno and Garnsey, 2008; for Korean, see Kwon et al., 2010). In addition, it is reported in studies on second language (L2) comprehension (Gass, 1979; Doughty, 1991; Hamilton, 1994). The findings of these studies have led to the conclusion that SRCs are easier to process cross-linguistically than ORCs both in L1 and L2 sentence processing. However, conflicting results have been reported on the processing-difficulty contrast between SRCs and ORCs in Mandarin Chinese (Chinese, hereafter; Chen et al., 2012). Reports of the ORC processing preference in Chinese<sup>1</sup> (e.g., Hsiao and Gibson, 2003; Hsu and Chen, 2007; Lin and Garnsey, 2011; Gibson and Wu, 2013; Sung et al., 2015) have posited a challenge to the presence of a universal SRC processing preference.

Previous studies (e.g., Hsiao and Gibson, 2003; Gibson and Wu, 2013) were conducted using self-paced reading tasks. This method usually requires readers to press the button for the occurrence of each word, which causes repeated interruptions in reading. The self-paced reading task cannot completely record or reflect the normal reading process, wherein readers

can move back and forth within a sentence, such as using regression or saccades. It also limits the scope of research and lacks certain online processing information. In particular, reading RC sentences requires dependency between the gap and filler so readers may read back and forth for integration and comprehension, which can be obtained with an eye-tracking device.

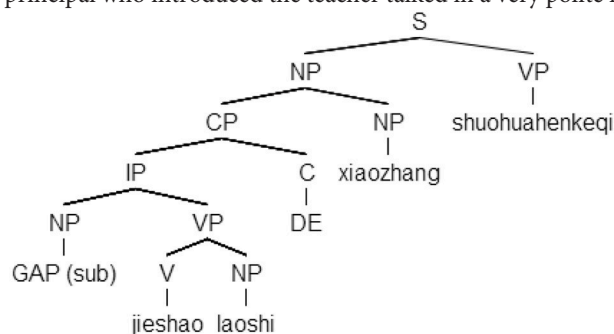
The goal of this study is to re-examine the L1 processing of Chinese RCs and further expand it to L2 RC processing. Since Chinese is a head-initial language with a head-final RC pattern while Japanese is a head-final language and the two languages share some syntactic similarities, such as RC location (both prenominal) and gap position (both prenominal), it would be interesting to see how the Japanese speakers, whose language has a typologically different RC structure, process Chinese RCs. Specifically, the study intends to see whether the Chinese syntactic nature or the universal SRC preference plays a more crucial role in the L2 RC processing by Japanese speakers. The native speakers of Chinese were also manipulated in the experiment. This study would like to serve as a pioneer eye-tracking research on L2 RC processing, which invites further studies on L2 learners with different language profiles.

## Aspects of Chinese and Japanese RC Chinese RC Structure and its Processing

The structure of a Chinese RC is different from that of an English RC. In Chinese, an RC precedes the noun to which it is attached and is transformed by adding an RC relativization marker, *de*, instead of a relative pronoun in an English RC. The extracted object or subject falls in the clause-final position. Examples and structural representations of Chinese RCs are given in (2)<sup>2</sup>.

### (2a) Chinese SRC

∅ 介绍 老师 的 校长 说话 很 客气  
[e jieshao laoshi de]<sub>RC</sub> xiaozhang shuohua hen keqi  
GAP introduce teacher REL principal talk very polite  
“The principal who introduced the teacher talked in a very polite manner.”

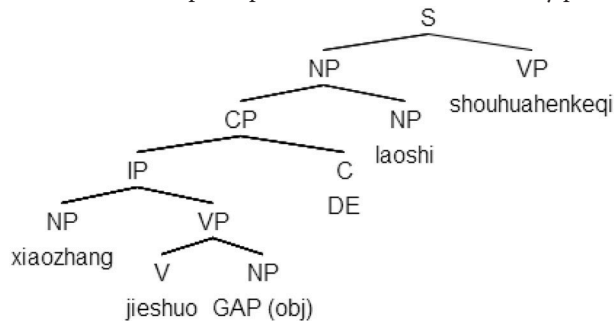


<sup>1</sup>Some studies on the processing of Basque RCs also showed an ORC preference (e.g., Carreiras et al., 2010), against the universal SRC preference.

<sup>2</sup>In (2), S, VP, CP, IP, C, and V are the subject, verb phrase, complementizer phrase, inflection phrase, complementizer, and verb, respectively.

## (2b) Chinese ORC

校長 介紹  $\emptyset$  的 老師 說話 很 客氣  
 [xiaozhang jieshao  $e$  de]RC laoshi shuohua hen keqi  
 principal introduce GAP REL teacher talk very polite  
 “The teacher whom the principal introduced talked in a very polite manner.”



In (2a), the noun *xiaozhang* “principal,” extracted from the subject position of the embedded verb *jieshao* “introduce,” serves as the head noun of the RC introduced by the relativizer *de*. *De* is treated as a relativization marker and is considered as a complementizer in phrasal structure (Aoun and Li, 2003). The extracted noun *xiaozhang* is coindexed with the gap and fills that gap. In (2b), the noun *laoshi* “teacher” is extracted from the object position and therefore, forms an ORC. It is noted that Chinese RCs are prenominal, which means they precede head nouns.

The processing of Chinese RCs has been intensively investigated over the past few years. Finding both preference types in Chinese indicates that the universal SRC preference is not consistent across languages. Therefore, the presence of processing asymmetry in Chinese RCs raises the issues of whether the processing pattern of Chinese RCs is language-specific, and whether discordant findings are related to the syntactically mixed patterns in Chinese RCs.

The processing of Chinese RCs has been intensively investigated using different methods. Among these studies, SRC preference has been reported in self-paced reading tasks (e.g., Lin and Bever, 2006a,b, 2007, 2011), in computational modeling (Chen et al., 2012), and in the relative frequency of occurrence in the corpus (Vasishth et al., 2013). Among those studies supporting a universal SRC preference, Lin and Bever (2006a,b, 2007, 2011) demonstrated that SRCs are easier than ORCs in Chinese. They examined how readers process RCs from two perspectives: RC modification (subject-modifying RC vs. object-modifying RC) and RC embeddedness (singly embedded RCs vs. doubly embedded RCs). They found that the reading times on both the relativizer and the head noun were significantly shorter for SRCs than for ORCs, irrespective of whether the RC modifies the subject or the object of the main clause. Their results suggest an SRC preference in Chinese, which is in line with the findings across languages.

However, the finding that SRCs are easier to process in Chinese has been challenged by other reports of an ORC preference (e.g., in self-paced reading tasks, see Hsiao and

Gibson, 2003; Hsu and Chen, 2007; Chen et al., 2008; Lin and Garnsey, 2011; Gibson and Wu, 2013; for a Mandarin-speaking aphasia case study, see Su et al., 2007). Among these works, the most often discussed is that of Hsiao and Gibson (2003). Those authors conducted a self-paced reading task with singly embedded and doubly embedded RCs, and with an RC-modifying subject of the main clauses. They found that in doubly embedded RCs, the reading times on the head noun and the embedded verb were shorter for ORCs than for SRCs. They demonstrated a preference for ORC in Chinese, implying that the processing of RCs in Chinese is language-specific.

### Japanese RC Structure and Processing

Like Chinese RCs, Japanese RCs come before the head nouns they modify, and therefore Japanese exhibits a prenominal RC pattern. Examples of Japanese RCs are given in (3).

- (3a) Japanese SRC  
 $\emptyset$  giin- o hinanshita kisha  
 [ $e_i$  senator- ACC attacked]RC reporter  
 GAP FILLER  
 “the reporter (who) attacked the senator”
- (3b) Japanese ORC  
 giin- ga  $\emptyset$  hinanshita kisha  
 [senator NOM  $e_i$  attacked]RC reporter  
 GAP FILLER  
 “the reporter (who) the senator attacked”

Adapted from Ueno and Garnsey (2008)

Japanese employs a different strategy to construct RCs, called the case-marking system, using case makers *o* and *ga* to indicate the syntactic function of the noun modified by the RC. Japanese does not have an overt relativizer, whereas Chinese has the relativizer *de* in RC.

Unlike reports of Chinese RCs having either an SRC and ORC preference, most research on the processing of Japanese RCs has demonstrated a preference for SRCs by native Japanese speakers (Miyamoto and Nakamura, 2003; Ueno and Garnsey, 2008).

## Comparison of Chinese and Japanese Structures

L2 sentence processing involves the syntactic structures of both L1 and L2. To understand and isolate the factors that potentially influence the processing of Chinese RCs by Japanese CSL learners, it is necessary to compare the structural properties of both involved languages. The structures of RC in Japanese and Chinese vary in several ways, as summarized in **Table 1**.

Due to the syntactic divergence between the two languages, it is expected that the different head positions between Chinese and Japanese may have certain effects on the L2 processing of Chinese, and therefore the two groups are hypothesized to show different processing patterns. Otherwise, if the two groups show similar processing patterns, then it would imply that the unique pattern of Chinese RCs may play a more crucial role.

## Theoretical Accounts on L2 Sentence Processing

Regarding L2 sentence processing, three theoretical accounts have been proposed to explain the difference in processing patterns between SRCs and ORCs: Noun Phrase Accessibility Hierarchy (NPAH, Keenan and Comrie, 1977), Structural Distance Hypothesis (SDH), and Linear Distance Hypothesis (LDH).

### Noun Phrase Accessibility Hierarchy

Keenan and Comrie (1977) proposed a universal tendency, called the NPAH, which was derived from the observations of syntactic forms in a large number of languages. The NPAH ranks the accessibility of the syntactic positions in a sentence as follows: subject, direct object, indirect object, oblique object, possessor, and object of comparison. Accordingly, a language that can relativize a given position in the hierarchy can also relativize all of its antecedent positions. Some studies suggest that the difficulty processing RCs experienced by L2 learners is associated with the NPAH (Gass, 1979, 1982; Pavesi, 1986; Doughty, 1988, 1991; Eckman et al., 1988; Wolfe-Quintero, 1992; Izumi, 2003). The findings in those studies parallel the typological implications captured by the NPAH. The NPAH hypothesizes that the degree of accessibility to RC formation across languages, and such a universal tendency, implies that SRCs make sentence processing easier than ORCs.

## Structural Distance Hypothesis

The SDH (Collins, 1994; Hamilton, 1995; O'Grady, 1997, 2001; O'Grady et al., 2003) states that the difficulty of an RC is determined by the depth of the gap corresponding to the relativized elements, and it is measured by counting the nodes between the gap and the filler of the RC.

In SRCs (e.g., 1a) the gap contains two phrasal nodes (IP and CP), while in ORCs (1b) the gap contains three phrasal nodes (VP, IP, and CP). Thus, the structural distance between the gap and its filler is greater in ORCs (1b) than in SRCs (1a).

## Linear Distance Hypothesis

The LDH (Tarallo and Myhill, 1983; Hawkins, 1989; O'Grady, 2001) presents a more straightforward measurement for the gap-filler distance: counting the intervening elements (words or words with discourse referents) between the gap and filler. In SRCs (e.g., 1a), there are three words and one discourse referent intervening between the gap and its filler, and in ORCs (e.g., 1b) there is only one word along the same path. Hence, the linear distance between gap and filler is greater in SRCs (1a) than in ORCs (1b).

## L2 Processing of RC

Regarding the processing of RCs by L2 learners, a great deal of previous studies have shown that L2 performance correlated with the prediction of NPAH (Gass, 1979, 1980, 1982; Hyltenstam, 1984; Pavesi, 1986; Doughty, 1988, 1991; Eckman et al., 1988; Wolfe-Quintero, 1992 among others). Those studies tested L2 learners with different L1 backgrounds through various tasks, including written sentence combination (Gass, 1979, 1980, 1982; Eckman et al., 1988), oral picture-cued production task (Pavesi, 1986), and guided oral production task (Wolfe-Quintero, 1992). The findings demonstrated that the participants performed better in SRCs than ORCs, implying that NPAH may still hold for the L2 sentence processing. Research on L2 processing in East Asian RCs, however, has arisen controversy to the acquisition difficulties across RC types (e.g., O'Grady et al., 2003; Jeon and Kim, 2007; Ozeki and Shirai, 2007; Yip and Matthews, 2007; Packard, 2008; Cui, 2013; Xu, 2014).

Several studies on the L2 processing of Chinese RC has reported inconsistent results with the acquisition difficulty/hierarchy made by NPAH. The results from those studies suggested that no clear processing asymmetry has been settled yet. First, Packard (2008) employed a self-paced reading task with both subject- and object-modifying SRCs and ORCs. He found that the participants read ORCs more quickly than SRCs. Moreover, Cui (2013) used a questionnaire and an online self-paced reading task to compare Chinese RC processing by L1 and L2 speakers. In her study, 24 native speakers of Chinese and 33 Chinese L2 learners (17 from head-initial L1 backgrounds and 16 from head-final L1 backgrounds) were recruited. The results of the questionnaire indicated that both L1 and L2 speakers found ORCs easier than SRCs. The data from the reading task showed that for L1 speakers, ORCs were read more quickly than SRCs only in subject-modifying RCs; for L2 speakers, no preference was found in the overall results, but an SRC preference was in the head-initial group. In addition, Xu

**TABLE 1 | Comparison of syntactic properties between Japanese and Chinese.**

Syntactic property	Chinese	Japanese
Canonical word order	S-V-O	S-O-V
Head position	Head-initial	Head-final
RC position	Prenominal	Prenominal
Relativizer	Yes (RC marker: DE)	No (by case-making system)
Gap position	Before the HN	Before the HN

(2014) conducted a written sentence combination task, testing 45 native English-speaking learners of Chinese on the production difficulty of four RC types: SRC, direct ORC, indirect ORC, object of preposition RC. She concluded that the participants' production difficulty fully follows the accessibility order of NPAH.

In a nutshell, the NPAH is typologically-driven and well-attested in many studies on L2 RC processing but it is still disputed for Chinese RCs. One may wonder if RC processing asymmetry is language- universal or specific in Chinese.

## THE CURRENT STUDY

This study focused on three major questions related to the conflicting research findings regarding processing of the two kinds of RCs:

1. Do speakers of Japanese (a head-final language) process Chinese RCs in a way similar to native speakers of Chinese (a head-initial language)?
2. How does the syntactic structure of L1 play a role in L2 processing of Chinese RCs?
3. Which of the theoretical hypotheses provides a better explanation of the results?

An eye-movement monitoring paradigm was employed to explore the reading patterns of Chinese RCs by Japanese CSL learners. The eye-movement technique enables us to obtain online information of consecutive reading as well as regression crucially relevant to RC processing. The eye-tracking indicators can be used to exam processing preference since fixation duration and the frequency of regressions increase as sentence becomes conceptually more difficult (Rayner, 1998).

## Interest Areas and Hypotheses

The interest areas, based on previous research, included the head noun and the embedded verb. These areas were examined in order to identify which type of RC is easier to process and where the processing difficulties (if any) arise. The head noun is of interest because it is the element that is extracted from the clause that later is transformed into an RC by adding the RC marker DE. In addition, head noun contrasts in the two RC types in terms of the syntactic functions (subject vs. object). The embedded verb was examined for two reasons: (1) the embedded verb syntactically governs the head noun of the RC, and (2) since Japanese is a head-final language, readers may pay special attention to the embedded verb, which may lead to different reading patterns from those of Chinese speakers.

The predictions of different theories regarding RCs, the head noun, and the embedded verb vary according to the factors emphasized, as follows:

1. The NPAH would predict an SRC preference because the subject is universally easier to relativize than the object, and participants would spend less time on the head noun of an SRC than of an ORC. That is, the processing time for an head noun would be *shorter* for an SRC than for an ORC. Although in the current study, all RCs modify the subjects of main

clauses (subject-modifying RCs), within the RC structure, we compared the processing difference between SRCs and ORCs. Then, NPAH would favor SRCs. The NPAH assumption would lead to another prediction: that L1 syntax does not influence L2 sentence processing.

2. The SDH would predict an SRC preference in Chinese in that the structural distance between gap and filler (i.e., the depth of embeddedness of the gap) is shorter in a SRC than in a ORC.
3. The LDH would predict an ORC preference in that the linear distance between the gap and its filler is shorter in a Chinese ORC than in a Chinese SRC.

The NPAH and SDH make the same prediction in Chinese RCs in terms of processing asymmetry. However, it should be noted that two accounts were based on different theoretical implications, the former came from observations of typologically different languages while the latter was built upon syntactic structures.

## METHOD

### Participants

Thirty-six native Japanese speakers were recruited at the Mandarin Training Center at National Taiwan Normal University. These participants had learned Chinese for a mean of 2.1 years, and their level of proficiency in the language corresponded to A1-B1 level according to their learning materials associated with the standard of the Common European Framework of Reference for Languages (CEFR). They came from classes of the same proficiency level. Thirty-eight native speakers of Mandarin Chinese who were college students were also recruited. The data of participants with a comprehension accuracy of <70% in the reading comprehension test were removed from the analysis. Based on the inclusion criteria, 33 native Japanese speakers and 38 native Chinese speakers were considered to be valid samples (11 male and 22 female Japanese, and 7 male, and 31 female Chinese). The native Japanese and Chinese speakers ranged in age from 21 to 50 years ( $M = 28.76$  years) and 20 to 49 years ( $M = 22.42$  years), respectively. All had normal or corrected-to-normal vision.

### Apparatus

The sentences were presented in black against a light-gray background on a 19-inch CHIMEI CMV A902 LCD monitor (1024 × 768-pixel resolution). Eye movements were recorded with an EyeLink 1000 eye tracker (SR Research, Canada). The sampling rate was set to 1000 Hz. The equipment comprised two personal computers (PCs) with Intel Core i5 3.2-GHz processors: one was a display PC that was responsible for presenting stimuli and controlling the experiment, and the other was a host PC that was responsible for monitoring and collecting eye-movement data. Participants were instructed to rest their head on a chinrest to minimize head movements. Although, viewing was performed binocularly, only data for the left eye were recorded. The programming was conducted using Experiment Builder 1.10.1 (SR Research Ltd. 2004–2010), and data were analyzed using Data Viewer 1.11.1 (SR Research Ltd. 2002–2011) and SPSS 18.

## Materials and Design

The experiment had a  $2 \times 2$  within-subject design. The independent variables were clause type (SRC vs. ORC) and distance (long vs. short), where distance was defined as the length between the gap and the head noun with which it was associated (long distance: 6–10 characters; short distance: 1–5 characters). The distance between the two syntactic dependents was manipulated by the additional modifiers preceding the head noun. The dependent variables were the accuracy rate in the reading comprehension test, reading time, gaze duration, regression-path duration, and regression rate.

The eye-movement task comprised 120 sentences: 60 experimental sentences (sentences with RCs) and 60 fillers (sentences without RCs). The experimental sentences comprised 30 SRCs and 30 ORCs, with 15 long-distance and 15 short-distance sentences in each. The frequencies of words in the

experimental sentences were calculated based on the frequency list of 8000 Chinese words compiled by the Steering Committee for the Test of Proficiency (SC-TOP) in Huayu, Taiwan, to ensure that (a) reading could be performed without vocabulary difficulties and (b) all the experimental sentences comprised phrases with similar difficulties for L2 learners, with only the complexities in the RCs varying. The average frequency of head nouns in SRCs was 0.014 ( $SD = 0.015$ ) and that in ORCs was 0.019 ( $SD = 0.029$ ). The difference between average frequencies of head nouns in SRCs and ORCs was not significant [ $t_{(14)} = 1.03, p = 0.32$ ]. The words used in the experimental sentences included the vocabulary at Level 1, 2, and 3 (corresponding to A1, A2, and B1 under CEFR) in Chinese 8000 words (SC-TOP). Sample stimuli are given in (4), where the head nouns and embedded verbs are bolded.

### (4a) Short-distance SRC

介紹 老師 的 校長 說話 很 客氣  
**jieshao** laoshi de **xiaozhang** shouhua hen keqi  
 introduce teacher REL principal talk very politely  
 “The principal who introduced the teacher talked in a very polite manner.”

### (4b) Short-distance ORC

校長 介紹 的 老師 說話 很 客氣  
 xiaozhang **jieshao** de **laoshi** shouhua hen keqi  
 principal introduce REL teacher talk very politely  
 “The teacher who the principal introduced talked in a very polite manner.”

### (4c) Long-distance SRC

介紹 上課 認真的 老師 的 校長 說話 很 客氣  
**jieshao** shangke-renzhen-de laoshi de **xiaozhang** shouhua hen keqi  
 introduce seriously-teaching teacher REL principal talk very politely  
 “The principal who introduced the hard-working teacher talked in a very polite manner.”

### (4d) Long-distance ORC

校長 介紹 的 上課 認真的 老師 說話 很 客氣  
 xiaozhang **jieshao** de shangke-renzhen-de **laoshi** shouhua hen keqi  
 principal introduce REL seriously-teaching teacher talk very politely  
 “The hard-working teacher who the principal introduced talked in a very polite manner.”

All experimental sentences and fillers were displayed in a single line in the middle of an LCD screen. The lengths of the sentences ranged from 12 to 21 characters, and the size of the characters was  $36 \times 36$  pixels, with intercharacter spaces of  $10 \times 36$  pixels. Each of the sentences was presented horizontally from left to right on the screen. The fillers covered various sentence structures, such as ba-construction and bei-construction. The fillers varied in length since the critical sentences included long-distance and short-distance sentences.

## Data Analysis

The data analysis included the accuracy of the reading comprehension test, reading time of full sentences, and eye-movement data. The eye-movement data were measured according to the gaze duration, regression-path duration, and regression rate. The gaze duration is the sum of all first-pass fixations on a region before the eyes move out of the region to either the right or left (Rayner, 1998). The regression-path

duration is the total time spent fixating on all of the target and pretarget regions, from the first fixation on a target region to fixation to the right of the target region (Rayner and Duffy, 1986; Liversedge et al., 1998). The regression rate is corresponding to the probability of rereading the target (Yen et al., 2008), i.e., the probability of regressions back into the target region after it has already been read. Fixations of  $<80$  or  $>1200$  ms (representing 4.15 and 5.00% of Japanese and Chinese language groups, respectively) were eliminated from the analyses (Liversedge et al., 2004; Drieghe et al., 2008; White, 2008; Slattery et al., 2013; Stites et al., 2013). Two-way repeated-measures ANOVAs were conducted for participants ( $F_1$ ) and items ( $F_2$ ). The analysis of eye-movement data was based on the trials in which no errors were made in the reading comprehension test.

## Procedure

This study was approved by National Science council in Taiwan with written informed consent from all participants.

The participants were asked to provide background information before they proceeded to begin doing the experiment. The eye-movement task, which took about 25 min for L1 participants and 40 min for L2 participants to complete, was then conducted. Each participant sat 70 cm in front of a screen, with his or her head leaning on a chinrest. At this viewing distance, each character subtended a visual angle of  $1.06^\circ$ . The task began with a 13-point calibration, followed by five practice trials. The practice trials were in the same format as the experimental trials. For drift correction, the participants were instructed to look at a dot positioned at the location of the first character of the sentence. After the participants fixated on the dot, the experimenter pressed a button and the sentence appeared on the screen. The participants viewed each complete sentence on the screen one at a time, and were instructed to read each sentence at their own pace. Their eye movements were tracked while reading the sentence. When the participants were finished reading, they were asked to answer a true-or-false reading comprehension question to ensure that they had understood the sentence. Once the reading comprehension question was presented, the subjects were unable to go back to the test sentence. There were 120 questions in total, with a 5-min break every 40 trials. The 13-point calibration was readministered after each break. The sentences were presented in a randomized order, meaning that each participant viewed them in a different order.

## RESULTS

### Accuracy of the Reading Comprehension Test

The means and standard deviations for accuracy of the comprehension test are given in **Table 2**.

For native Chinese speakers, the main effect of clause type on accuracy was marginally significant [ $F_{1(1, 37)} = 3.91$ ,  $MSE = 0.001$ ,  $p = 0.055$ , partial eta squared ( $\eta^2$ ) = 0.10;  $F_{2(1, 14)} = 0.96$ ,  $MSE = 0.002$ ,  $p = 0.34$ ,  $\eta^2 = 0.08$ ]. The accuracy was higher for SRCs ( $M = 0.980$ ,  $SD = 0.055$ ) than for ORCs ( $M = 0.969$ ,  $SD = 0.056$ ). However, there was no significant main effect of distance or significant interaction of clause type and distance (all  $ps > 0.50$ ).

For Japanese speakers, the main effect of distance on accuracy was significant in the analysis by participants [ $F_{1(1, 32)} = 4.15$ ,  $MSE = 0.006$ ,  $p < 0.05$ ,  $\eta^2 = 0.11$ ;  $F_{2(1, 14)} = 0.05$ ,  $MSE = 0.019$ ,  $p = 0.83$ ,  $\eta^2 = 0.00$ ]. The accuracy was higher for short-distance sentences ( $M = 0.905$ ,  $SD = 0.126$ ) than for long-distance sentences ( $M = 0.877$ ,  $SD = 0.137$ ). However, there was no significant main effect of clause type or significant interaction of clause type and distance (all  $p > 0.10$ ).

### Reading Time of Full Sentences

The means and standard deviations for reading time of full sentences are given in **Table 3**.

For native Chinese speakers, the main effect of clause type on reading time was significant [ $F_{1(1, 37)} = 16.50$ ,  $MSE = 0.233$ ,  $p < 0.001$ ,  $\eta^2 = 0.31$ ;  $F_{2(1, 14)} = 11.61$ ,  $MSE = 0.136$ ,  $p < 0.01$ ,  $\eta^2 = 0.45$ ]. The reading time was significantly longer for full sentences in SRCs ( $M = 4.05$  s,  $SD = 1.36$  s) than for those in ORCs ( $M = 3.74$  s,  $SD = 1.29$  s). The main effect of distance on reading time was significant [ $F_{1(1, 37)} = 181.89$ ,  $MSE = 0.507$ ,  $p < 0.001$ ,  $\eta^2 = 0.83$ ;  $F_{2(1, 14)} = 449.28$ ,  $MSE = 0.077$ ,  $p < 0.001$ ,  $\eta^2 = 0.97$ ]. The reading time was significantly longer for long-distance sentences ( $M = 4.67$  s,  $SD = 1.31$  s) than for short-distance sentences ( $M = 3.12$  s,  $SD = 0.80$  s). The interaction between clause type and distance was also significant [ $F_{1(1, 37)} = 10.41$ ,  $MSE = 0.298$ ,  $p < 0.01$ ,  $\eta^2 = 0.22$ ;  $F_{2(1, 14)} = 12.00$ ,  $MSE = 0.120$ ,  $p < 0.01$ ,  $\eta^2 =$

**TABLE 2 | Mean (M) and standard deviation (SD) values for the accuracy of the post-stimulus reading comprehension test.**

Type	Chinese				Japanese			
	Long		Short		Long		Short	
	M	SD	M	SD	M	SD	M	SD
ORC	0.970	0.057	0.968	0.055	0.889	0.093	0.884	0.153
SRC	0.977	0.053	0.982	0.056	0.865	0.169	0.925	0.087

Long and short refer to the distance between the gap and the HN.

**TABLE 3 | Mean (M) and standard deviation (SD) values for Reading time of full sentences (sec).**

Type	Chinese				Japanese			
	Long		Short		Long		Short	
	M	SD	M	SD	M	SD	M	SD
ORC	4.37	1.30	3.10	0.90	8.13	1.41	5.59	1.03
SRC	4.98	1.24	3.13	0.68	9.14	1.81	6.13	1.15

Long and short refer to the distance between the gap and the HN.



0.46]. A simple main effect test showed that the reading time was significantly longer for full sentences in long-distance SRCs ( $M = 4.94$  s,  $SD = 0.64$  s) than for those in long-distance ORCs ( $M = 4.31$  s,  $SD = 0.33$  ms) [ $F_{1(1, 74)} = 26.08$ ,  $MSE = 0.265$ ,  $p < 0.001$ ,  $\eta^2 = 0.26$ ;  $F_{2(1, 28)} = 23.58$ ,  $MSE = 0.128$ ,  $p < 0.001$ ,  $\eta^2 = 0.46$ ]. Regardless of clause type, the reading time was significantly longer for long-distance sentences than for short-distance sentences (all  $ps < 0.001$ ; see **Figure 1**).

For Japanese speakers, the main effect of clause type on reading time was significant [ $F_{1(1, 32)} = 41.54$ ,  $MSE = 0.481$ ,  $p < 0.001$ ,  $\eta^2 = 0.56$ ;  $F_{2(1, 14)} = 24.59$ ,  $MSE = 0.399$ ,  $p < 0.001$ ,  $\eta^2 = 0.64$ ]. The reading time was significantly longer for sentences in SRCs ( $M = 7.64$  s,  $SD = 2.13$  s) than for those in

ORCs ( $M = 6.86$  s,  $SD = 1.77$  s). The main effect of distance on reading time was significant [ $F_{1(1, 32)} = 203.67$ ,  $MSE = 1.244$ ,  $p < 0.001$ ,  $\eta^2 = 0.86$ ;  $F_{2(1, 14)} = 324.52$ ,  $MSE = 0.370$ ,  $p < 0.001$ ,  $\eta^2 = 0.96$ ]. The reading time was significantly longer for long-distance sentences ( $M = 8.63$  s,  $SD = 1.70$  s) than for short-distance sentences ( $M = 5.86$  s,  $SD = 1.13$  s). However, there was no significant interaction of clause type and distance (both  $ps > 0.06$ ).

## Eye-Movement Data

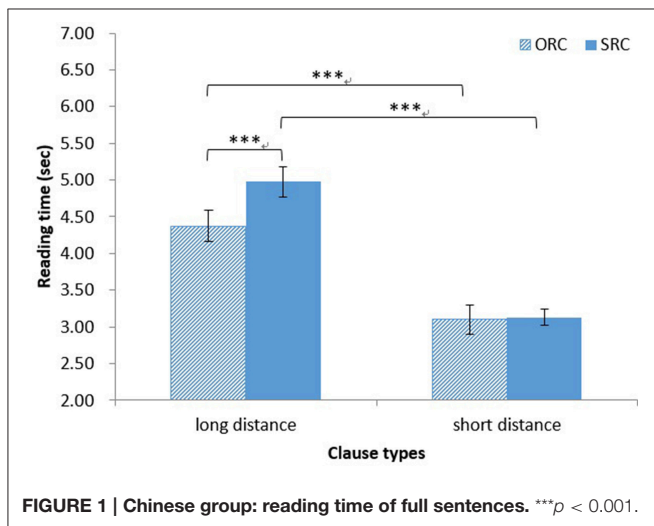
**Table 4** lists the descriptive statistics for the gaze duration, regression-path duration, and regression rate. These indices are discussed in detail below.

### Chinese Speakers

#### Head nouns

Head nouns were measured using gaze duration, regression-path duration, and regression rate.

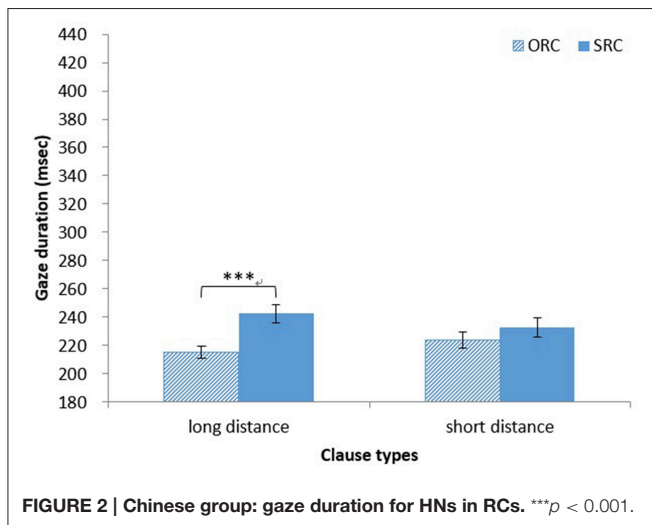
**Gaze duration.** The main effect of clause type on gaze duration was significant [ $F_{1(1, 37)} = 16.75$ ,  $MSE = 744$ ,  $p < 0.001$ ,  $\eta^2 = 0.31$ ;  $F_{2(1, 14)} = 12.08$ ,  $MSE = 390$ ,  $p < 0.01$ ,  $\eta^2 = 0.46$ ]. The gaze duration was significantly longer for head nouns in SRCs ( $M = 238$  ms,  $SD = 41$  ms) than for those in ORCs ( $M = 220$  ms,  $SD = 31$  ms). The main effect of distance was not significant (both  $ps > 0.90$ ). The interaction between clause type and distance was also significant in the analysis by participants [ $F_{1(1, 37)} = 5.64$ ,  $MSE = 570$ ,  $p < 0.05$ ,  $\eta^2 = 0.13$ ;  $F_{2(1, 14)} = 3.59$ ,  $MSE = 306$ ,  $p = 0.08$ ,  $\eta^2 = 0.20$ ]. A simple main effect test showed that the gaze duration was significantly longer for Head nouns in long-distance SRCs ( $M = 243$  ms,  $SD = 39$  ms) than for those in long-distance ORCs ( $M = 215$  ms,  $SD = 26$  ms)



**TABLE 4 | Data for eye-movement indices for each critical region.**

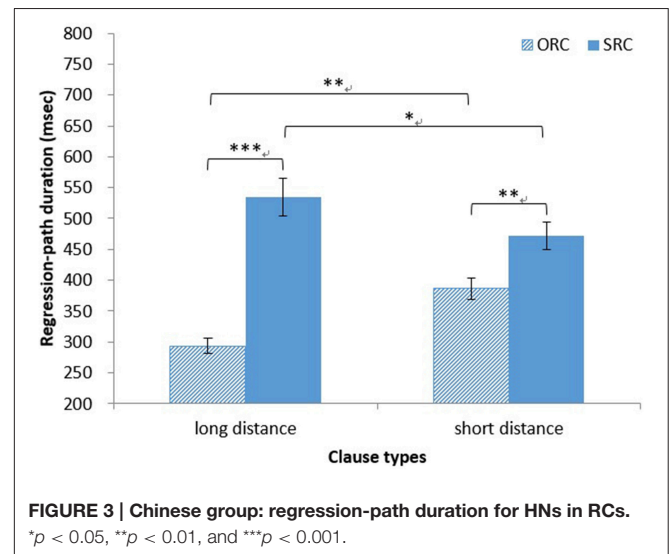
	Chinese				Japanese			
	HN		EV		HN		EV	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<b>GAZE DURATION (ms)</b>								
ORC-L	215	26	277	51	341	65	345	67
ORC-S	224	34	272	51	371	76	346	78
SRC-L	243	39	241	57	370	70	414	139
SRC-S	233	42	231	52	364	83	397	123
<b>REGRESSION-PATH DURATION (ms)</b>								
ORC-L	294	72	N/A		434	133	N/A	
ORC-S	387	105	N/A		512	160	N/A	
SRC-L	535	184	N/A		642	293	N/A	
SRC-S	472	134	N/A		564	197	N/A	
<b>REGRESSION RATE</b>								
ORC-L	0.32	0.15	0.66	0.21	0.39	0.17	0.69	0.20
ORC-S	0.35	0.19	0.70	0.18	0.43	0.17	0.71	0.18
SRC-L	0.28	0.17	0.70	0.28	0.38	0.18	0.78	0.17
SRC-S	0.32	0.15	0.62	0.30	0.36	0.18	0.78	0.19

L, long distance between the gap and the HN; S, short distance between the gap and the HN; N/A, not acquired.



$[F_{1(1, 74)} = 21.56, MSE = 657, p < 0.001, \eta^2 = 0.23; F_{2(1, 28)} = 14.88, MSE = 348, p < 0.001, \eta^2 = 0.35$ ; see **Figure 2**].

**Regression-path duration.** The main effect of clause type on regression-path duration was significant [ $F_{1(1, 37)} = 69.01, MSE = 14, 662, p < 0.001, \eta^2 = 0.65; F_{2(1, 14)} = 56.97, MSE = 6828, p < 0.001, \eta^2 = 0.80$ ]. The regression-path duration was significantly longer for head nouns in SRCs ( $M = 503$  ms,  $SD = 164$  ms) than for those in ORCs ( $M = 340$  ms,  $SD = 101$  ms). The main effect of distance was not significant (both  $ps > 0.30$ ). The interaction between clause type and distance was also significant [ $F_{1(1, 37)} = 21.85, MSE = 10, 461, p < 0.001, \eta^2 = 0.37; F_{2(1, 14)} = 38.42, MSE = 2629, p < 0.001, \eta^2 = 0.73$ ]. A simple main effect test showed that the regression-path duration was significantly longer for head nouns in long-distance SRCs ( $M = 535$  ms,  $SD = 184$  ms) than for those in long-distance ORCs ( $M = 294$  ms,  $SD = 72$  ms) [ $F_{1(1, 74)} = 87.66, MSE = 12, 561, p < 0.001, \eta^2 = 0.54; F_{2(1, 28)} = 93.73, MSE = 4728, p < 0.001, \eta^2 = 0.77$ ]. The regression-path duration was significantly longer for head nouns in short-distance SRCs ( $M = 472$  ms,  $SD = 134$  ms) than for those in short-distance ORCs ( $M = 387$  ms,  $SD = 105$  ms) [ $F_{1(1, 74)} = 11.09, MSE = 12, 561, p < 0.01, \eta^2 = 0.13; F_{2(1, 28)} = 9.90, MSE = 4728, p < 0.01, \eta^2 = 0.26$ ]. The regression-path duration was significantly longer for head nouns in long-distance SRCs ( $M = 535$  ms,  $SD = 184$  ms) than for those in short-distance SRCs by items ( $M = 472$  ms,  $SD = 134$  ms) [ $F_{1(1, 74)} = 5.15, MSE = 14, 339, p = 0.03, \eta^2 = 0.07; F_{2(1, 28)} = 10.35, MSE = 3328, p < 0.01, \eta^2 = 0.27$ ]. The regression-path duration was significantly longer for head nouns in short-distance ORCs ( $M = 387$  ms,  $SD = 105$  ms) than for those in long-distance ORCs ( $M = 294$  ms,  $SD = 72$  ms) [ $F_{1(1, 74)} = 11.41, MSE = 14, 339, p < 0.01, \eta^2 = 0.13; F_{2(1, 28)} = 20.92, MSE = 3328, p < 0.001, \eta^2 = 0.43$ ]. Regardless of the sentence distance, the regression-path duration was significantly longer for head nouns in SRCs than for those in ORCs (all  $ps < 0.01$ ; see **Figure 3**).



**Regression rate.** The main effect of clause type on regression rate was not significant (both  $ps > 0.10$ ), as were the main effect of distance (both  $ps > 0.05$ ) and the interaction between clause type and distance (both  $ps > 0.80$ ).

The results for head nouns showed that gaze duration and regression-path duration on SRCs were longer than those on ORCs. Hence, the results generally suggested an ORC preference for Chinese speakers.

#### Embedded verbs

Embedded verbs were measured using gaze duration and regression rate. Note that the measure of regression-path duration was not used since the positions of the embedded verbs in the two types of RCs are different. In particular, the embedded verb of SRC is in sentence-initial position so that its regression time may be underestimated. Thus, the indicator regression-path duration involving regression time was excluded for the analysis of embedded verbs.

**Gaze duration.** The main effect of clause type on gaze duration was significant [ $F_{1(1, 37)} = 19.98, MSE = 2819, p < 0.001, \eta^2 = 0.35; F_{2(1, 14)} = 84.35, MSE = 202, p < 0.001, \eta^2 = 0.86$ ]. The gaze duration was significantly longer for embedded verbs in ORCs ( $M = 274$  ms,  $SD = 51$  ms) than for those in SRCs ( $M = 236$  ms,  $SD = 55$  ms). However, there was no significant main effect of distance or significant interaction of clause type and distance (all  $ps > 0.10$ ).

**Regression rate.** The main effect of clause type on regression rate was not significant (both  $ps > 0.10$ ), as was the main effect of distance (both  $ps > 0.30$ ). However, the interaction between clause type and distance was significant [ $F_{1(1, 37)} = 6.75, MSE = 0.018, p < 0.05, \eta^2 = 0.15; F_{2(1, 14)} = 4.59, MSE = 0.005, p = 0.05, \eta^2 = 0.25$ ]. A simple main effect test showed that the regression rate was marginally significantly longer for embedded verbs in long-distance SRCs ( $M = 0.70, SD = 0.28$ ) than for those in short-distance SRCs in the analysis by participants ( $M = 0.62, SD = 0.30$ ) [ $F_{1(1, 74)} = 5.54, MSE = 0.021, p = 0.02$ ,

$\eta^2 = 0.07$ ;  $F_{2(1, 28)} = 1.03$ ,  $MSE = 0.011$ ,  $p = 0.32$ ,  $\eta^2 = 0.04$ ; see **Figure 4**].

The results for embedded verbs showed that gaze duration on ORCs was longer than that on SRCs. The interpretation of such finding will be particularly discussed in Section ORC Processing Preference.

## Japanese Speakers

### Head nouns

Head nouns were measured using gaze duration, regression-path duration, and regression rate.

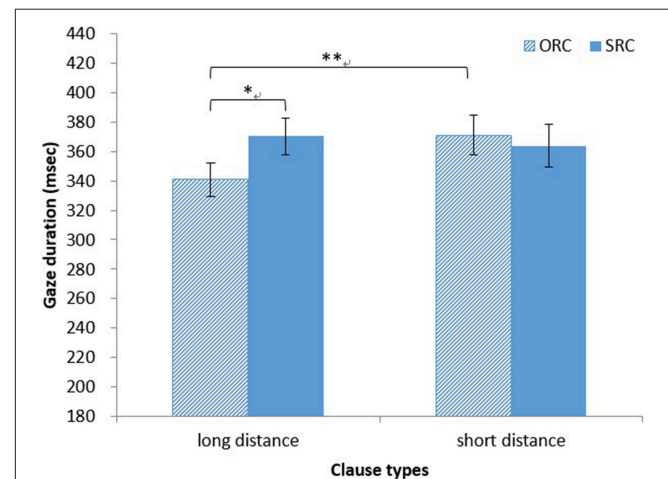
**Gaze duration.** The main effect of clause type on gaze duration was not significant (both  $ps > 0.20$ ), as was the main effect of distance (both  $ps > 0.05$ ). However, the interaction between clause type and distance was significant [ $F_{1(1, 32)} = 4.56$ ,  $MSE = 2393$ ,  $p < 0.05$ ,  $\eta^2 = 0.13$ ;  $F_{2(1, 14)} = 7.69$ ,  $MSE = 605$ ,  $p < 0.05$ ,  $\eta^2 = 0.35$ ]. A simple main effect test showed that the gaze duration was significantly longer for head nouns in long-distance SRCs ( $M = 370$  ms,  $SD = 70$  ms) than for those in long-distance ORCs ( $M = 341$  ms,  $SD = 65$  ms) [ $F_{1(1, 64)} = 5.86$ ,  $MSE = 2401$ ,  $p < 0.05$ ,  $\eta^2 = 0.08$ ;  $F_{2(1, 28)} = 5.29$ ,  $MSE = 1516$ ,  $p < 0.05$ ,  $\eta^2 = 0.16$ ]. The gaze duration was significantly longer for head nouns in short-distance ORCs ( $M = 371$  ms,  $SD = 76$  ms) than for those in long-distance ORCs [ $F_{1(1, 64)} = 7.26$ ,  $MSE = 2033$ ,  $p < 0.01$ ,  $\eta^2 = 0.10$ ;  $F_{2(1, 28)} = 10.94$ ,  $MSE = 568$ ,  $p < 0.01$ ,  $\eta^2 = 0.28$ ; see **Figure 5**].

**Regression-path duration.** The main effect of clause type on regression-path duration was significant [ $F_{1(1, 32)} = 20.14$ ,  $MSE = 27, 567$ ,  $p < 0.001$ ,  $\eta^2 = 0.39$ ;  $F_{2(1, 14)} = 20.43$ ,  $MSE = 14, 430$ ,  $p < 0.001$ ,  $\eta^2 = 0.59$ ]. The regression-path duration was significantly longer for head nouns in SRCs ( $M = 603$  ms,  $SD = 253$  ms) than for those in ORCs ( $M = 473$  ms,  $SD = 152$  ms). The main effect of distance was not significant (both  $ps > 0.90$ ). The interaction between clause type and distance was also significant [ $F_{1(1, 32)} = 8.10$ ,  $MSE = 24, 939$ ,  $p < 0.01$ ,  $\eta^2 = 0.20$ ;  $F_{2(1, 14)} = 12.80$ ,  $MSE = 7446$ ,  $p < 0.01$ ,  $\eta^2 = 0.48$ ]. A simple main effect test showed that the regression-path

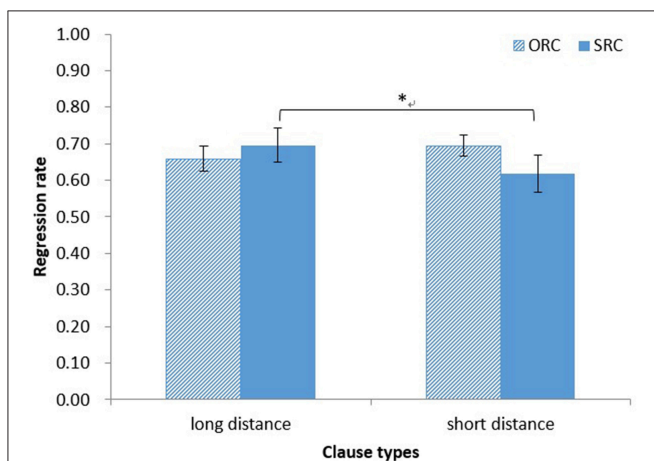
duration was significantly longer for head nouns in long-distance SRCs ( $M = 642$  ms,  $SD = 293$  ms) than for those in long-distance ORCs ( $M = 434$  ms,  $SD = 133$  ms; [ $F_{1(1, 64)} = 27.17$ ,  $MSE = 26, 253$ ,  $p < 0.001$ ,  $\eta^2 = 0.30$ ;  $F_{2(1, 28)} = 33.15$ ,  $MSE = 10, 938$ ,  $p < 0.001$ ,  $\eta^2 = 0.54$ ; see **Figure 6**].

**Regression rate.** The main effect of clause type on regression rate was not significant (both  $ps > 0.05$ ), as were the main effect of distance (both  $ps > 0.60$ ) and the interaction between clause type and distance (both  $ps > 0.10$ ).

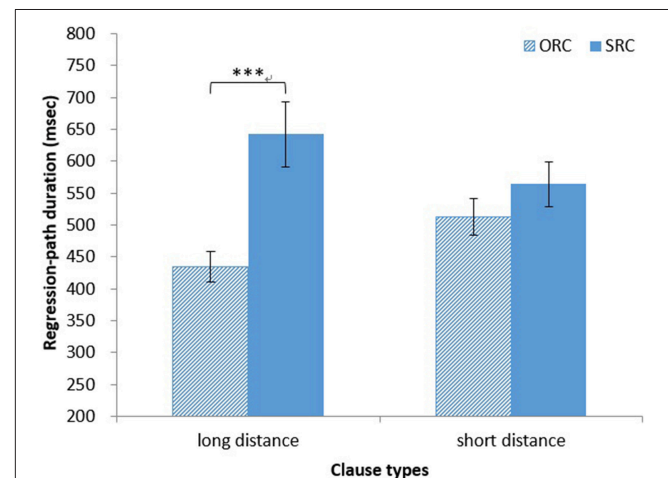
The results for head nouns showed that gaze duration and regression-path duration on SRCs were longer than those on ORCs. The results generally suggested an ORC preference for Japanese speakers.



**FIGURE 5 | Japanese group: gaze duration for HNs in RCs.** \* $p < 0.05$ , \*\* $p < 0.01$ .



**FIGURE 4 | Chinese group: regression rate for EVs in RCs.** \* $p < 0.05$ .



**FIGURE 6 | Japanese group: regression-path durations for HNs in RCs.** \*\*\* $p < 0.001$ .

### Embedded verbs

Embedded verbs were measured using gaze duration and regression rate.

**Gaze duration.** The main effect of clause type on gaze duration was significant [ $F_{1(1, 32)} = 8.23$ ,  $MSE = 14, 339$ ,  $p < 0.01$ ,  $\eta^2 = 0.20$ ;  $F_{2(1, 14)} = 15.87$ ,  $MSE = 2754$ ,  $p < 0.01$ ,  $\eta^2 = 0.53$ ]. The gaze duration was significantly longer for embedded verbs in SRCs ( $M = 406$  ms,  $SD = 132$  ms) than for those in ORCs ( $M = 346$  ms,  $SD = 73$  ms). However, there was no significant main effect of distance or significant interaction between clause type and distance (all  $ps > 0.30$ ).

**Regression rate.** The main effect of clause type on regression rate was significant [ $F_{1(1, 32)} = 6.80$ ,  $MSE = 0.033$ ,  $p < 0.05$ ,  $\eta^2 = 0.18$ ;  $F_{2(1, 14)} = 42.31$ ,  $MSE = 0.002$ ,  $p < 0.001$ ,  $\eta^2 = 0.75$ ]. The regression rate was significantly higher for embedded verbs in SRCs ( $M = 0.78$ ,  $SD = 0.18$ ) than for those in ORCs ( $M = 0.70$ ,  $SD = 0.19$ ). However, there was no significant main effect of distance or significant interaction between clause type and distance (all  $ps > 0.20$ ).

The results for embedded verbs showed that gaze duration on SRCs was longer than that on ORCs and regression rate for SRCs was higher than that for ORCs. The results generally suggested an ORC preference for Japanese speakers.

Taken together, the results comparing (1) SRC and ORC processing in the long-distance conditions and (2) SRC and ORC processing in the short-distance conditions are summarized in **Table 5**.

The overall results, as shown by the indicators in **Table 4**, suggested that the Japanese group revealed an ORC preference. Similarly, the Chinese group demonstrated an ORC preference, except for the gaze duration of embedded verbs, which will be discussed in next section.

## DISCUSSION

In this study we used an eye-tracking technique to explore the difficulty experienced by Japanese learners when they are processing Chinese RCs. Overall, the results showed that ORCs were easier to process for Japanese CSL learners, which is similar to the pattern exhibited by Chinese speakers in terms of processing asymmetry. Our findings were consistent with the predictions of accounts based on the LDH, but not those of the NPAH and SDH. In this section we discuss the findings under

the theoretical framework of RC processing in order to address our research questions, and then consider the implications for L2 sentence comprehension.

### ORC Processing Preference

The results generally suggest an ORC preference for both the Chinese speakers and Japanese CSL learners. For the Chinese speakers, an ORC preference was evident from the following three results: (1) the gaze duration for head nouns in ORCs was shorter than that in SRCs, (2) the gaze duration for head nouns in long-distance ORCs was shorter than that in long-distance SRCs, and (3) the regression-path duration for head nouns in ORCs was shorter than for those in SRCs. The two indicators—gaze duration and regression-path duration—can reflect the initial and later stages of sentence processing, respectively. On the other hand, although the gaze duration for embedded verbs in ORCs was longer than for those in SRCs, we noticed that the Chinese speakers skipped around 63.3% of the embedded verbs in SRCs. It seems that the components in the sentence-initial position tend to be skipped. The finding that the gaze duration for embedded verbs in ORCs was longer than for those in SRCs may be due to the skipping rate being higher for the sentence-initial embedded verbs in SRCs. Hence, the overall results of the eye-movement data from Chinese speakers indicate that ORCs were easier to process than SRCs.

As for the Japanese CSL learners, their processing pattern also demonstrated an ORC preference, which can be observed from the following five findings: (1) the gaze duration for head nouns in long-distance ORCs was shorter than that in long-distance SRCs, (2) the regression-path duration for head nouns in ORCs was shorter than for those in SRCs, (3) the regression-path duration for head nouns in long-distance ORCs was shorter than for those in long-distance SRCs, (4) the gaze duration for embedded verbs in ORCs was shorter than for those in SRCs, and (5) the regression rate for embedded verbs in ORCs was lower than for those in SRCs. These results showed that Japanese CSL learners spent a shorter time on head nouns in ORCs in the initial reading (as reflected by the shorter gaze duration) and regression process (as reflected by the shorter regression-path duration), and they spent a shorter time on embedded verbs in ORCs in the initial reading process (as reflected by the shorter gaze duration) as well as in the regression process (as reflected by the lower regression rate). Thus, these results suggest an ORC preference for the Japanese CSL learners.

**TABLE 5 | Summary Table: results of clause type, and clause type  $\times$  distance.**

	Chinese		Japanese	
	HN	EV	HN	EV
Gaze duration	ORC < SRC ORC-L < SRC-L	SRC < ORC	ORC-L < SRC-L	ORC < SRC
Regression-path duration	ORC < SRC	N/A	ORC < SRC ORC-L < SRC-L	N/A
Regression rate	No significant difference	##	No significant difference	ORC < SRC

The results comparing (1) short-distance and long-distance SRC processing and (2) short-distance and long-distance ORC processing are NOT presented here. The symbol ## indicates that the significant result here (SRC-L > SRC-S) is not relevant to the discussion on the interaction between clause type and distance.

In summary, the overall results from the two language groups suggest that the two groups exhibited a similar pattern in terms of processing asymmetry; that is, a tendency toward ORC preference. The finding of an ORC preference was inconsistent with the prediction of the NPAH, which proposes that it is easier to relativize the subject than the object across languages. Chinese RC processing seems to be language-specific. In addition, consider the two hypotheses (LDH and SDH), which focus on gap–filler dependencies in RCs. Specifically, the LDH proposes that the distance between a filler and its gap is determined by the linear/temporal distance, while the SDH emphasizes the role of the hierarchical phrase-structural distance between the filler and the gap. The predictions of the two hypotheses diverge in the processing asymmetry of Chinese RCs since the LDH and SDH use different methods to calculate the gap–filler distance. In other words, in the case of Chinese RC processing, the LDH would predict an ORC preference while the SDH would predict an SRC preference. Thus, our results provide solid evidence in support of the LDH.

## L1 and L2 Processing of Chinese RCs

In an attempt to further understand the relationship between L1 and L2 processing of Chinese RCs, we compared the processing patterns of Chinese speakers and Japanese CSL learners. In terms of processing asymmetry, both the Chinese and Japanese groups generally exhibited an ORC preference. This ORC preference in both the L1 and L2 groups supports the LDH, which indicates that the syntactic structure of Chinese plays a crucial role. On the one hand, the Chinese group exhibited an ORC preference, which can probably be attributed to the syntactic structure of RCs in Chinese, which is a head-initial language with a head-final RC construction. The tendency toward ORCs being easier to process implies that the processing difficulty can be reflected by the linear distance between the gap and the filler. In the structure of a Chinese RC, the linear distance between the gap and the filler is shorter in ORCs than in SRCs, and therefore the tendency toward an ORC preference suggests an apparent effect of linear distance on Chinese RC processing asymmetry. Our finding of an ORC preference for Chinese speakers concurs with those of previous studies (Hsiao and Gibson, 2003; Hsu and Chen, 2007; Lin and Garnsey, 2011; Gibson and Wu, 2013). On the other hand, our finding that the Japanese group also exhibited an ORC preference indicated that the syntactic structure of the target language (i.e., Chinese) was a determining factor in the RC processing. Recall that an SRC preference appears in the L1 processing of Japanese RC (Ueno and Garnsey, 2008), whereas an ORC preference was found in the L2 processing of Chinese RCs in the current study. If the NPAH—which proposes that the subject is easier to relativize than the object—holds in the RC processing of CSL, then Japanese learners will exhibit an SRC preference. Also, the SDH predicts that Japanese CSL learners will show an SRC preference since the structural distance between the gap and the filler in SRCs is shorter than in ORCs. However, the current results showed that Japanese CSL learners demonstrated an ORC preference in the processing of Chinese RCs, as did the Chinese speakers. The processing asymmetry exhibited by Japanese CSL learners was similar to that of Chinese speakers.

Thus, we argue that it is the syntactic structure of the target language (i.e., Chinese) that influences how RCs are processed by Japanese CSL learners.

## The Effect of Modifiers

It is worth mentioning the interesting finding that both the L1 and L2 groups spent less time on the head nouns in the long-distance ORCs than on those in the short-distance ORCs. For the L1 group, the regression-path duration for the head nouns in long-distance ORCs was shorter than for those in short-distance ORCs. Also, for the L2 group, the gaze duration for the head nouns in long-distance ORCs were shorter than for those in short-distance ORCs. Together these results indicate that it was easier to process the head nouns in long-distance ORCs than those in short-distance ORCs. This finding contrasts with our expectation that the head noun in a long sentence would take a longer time to process than that in a short sentence, because in the long-distance sentence there are more antecedent elements before the head noun, which is assumed to consume more cognitive resources. Counterintuitively, it was found that the head noun in the long-distance ORCs had a shorter processing time, and we speculate that this is because the modifiers for the head nouns in long-distance ORCs can provide information to help readers predict the upcoming head nouns. The modifiers of head nouns in long-distance ORCs seem to better facilitate the processing of head nouns. This finding therefore, suggests how modifiers influence sentence processing.

## CONCLUDING REMARKS

Most of the previous studies on Chinese RCs have focused on processing by L1 speakers, and few have considered the syntactic comprehension of L2 speakers. The current study employed an eye-movement technique to investigate the RC processing by Japanese CSL learners. The eye-movement data—reflecting both gaze duration and regression patterns—revealed that Japanese CSL learners and Chinese speakers have a tendency toward ORC preference. From a theoretical perspective, this ORC preference in processing Chinese RCs supports the prediction of the LDH that the key determinant of RC difficulty is the length of the gap from the head noun with which it is associated; that is, the linear distance between the gap and the filler. From an empirical perspective, these findings indicate that L1 and L2 speakers both demonstrate a tendency toward ORC preference, which suggests that the syntactic structure of RCs in Chinese is a dominating factor. The processing of Chinese RCs is language-specific in that it has a mixed pattern of a head-initial language with a head-final RC structure, constructing a shorter filler-gap distance in ORCs than that in SRCs, which perhaps makes ORCs easier to process than SRCs. In conclusion, this research expands our understanding of RC processing from L1 speakers to L2 learners of Chinese, and should provide a useful basis for further studies on the L2 processing of Chinese RCs with evidence from typologically different languages.

## ACKNOWLEDGMENTS

This research is partially supported by the “Aim for the Top University Project” and “Center of Learning Technology for Chinese” of National Taiwan Normal University (NTNU),

sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under MOST 104-2511-S-003 -012 -MY3; MOST 104-2511-S-003 -018 -MY3.

## REFERENCES

- Aoun, J., and Li, Y.-H. A. (2003). *Essays on the Representational and Derivational Nature of Grammar: The Diversity of wh-Constructions*. Cambridge: MIT Press.
- Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., and Laka, I. (2010). Subject relative clauses are not universally easier to process: evidence from Basque. *Cognition* 115, 79–92. doi: 10.1016/j.cognition.2009.11.012
- Chen, B., Ning, A., Bi, H., and Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychol.* 129, 61–65. doi: 10.1016/j.actpsy.2008.04.005
- Chen, Z., Grove, K., and Hale, J. (2012). “Structural expectations in Chinese relative clause comprehension,” in *Proceedings of the 29th West Coast Conference on Formal Linguistics (WCCFL29)*, (Somerville, MA: Cascadia Proceedings Project), 29–37.
- Collins, C. (1994). Economy of derivation and the generalized proper binding condition. *Linguist. Inquiry* 25, 45–61.
- Cui, Y. (2013). “L2 processing of relative clauses in Mandarin,” in *Arizona Working Papers in SLA and Teaching*, Vol. 20 (Tucson, AZ), 20–39.
- Doughty, C. (1991). Second language instruction does make a difference: evidence from an empirical study of SL relativization. *Stud. Sec. Lang. Acquisition* 13, 431–469. doi: 10.1017/S0272263100010287
- Doughty, C. (1988). *The Effects of Instruction on the Acquisition of Relativization in English as a Second Language*. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- Drieghe, D., Pollatsek, A., Staub, A., and Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *J. Exp. Psychol.* 34, 1552–1560. doi: 10.1037/a0013017
- Eckman, F., Bell, L., and Nelson, D. (1988). On the generalization of relative clause instruction in the acquisition of English as a Second Language. *Appl. Linguist.* 29, 1–20. doi: 10.1093/applin/9.1.1
- Frazier, L. (1987). Syntactic processing: evidence from Dutch. *Nat. Lang. Linguist. Theory* 5, 519–559. doi: 10.1007/BF00138988
- Gass, S. (1979). Language transfer and universal grammatical relations. *Lang. Learn.* 29, 327–344. doi: 10.1111/j.1467-1770.1979.tb01073.x
- Gass, S. (1980). “An investigation of syntactic transfer in adult second language learners,” in *Research in Second Acquisition: Selected Paper from the Los Angeles Second Language Acquisition Forum*, eds R. Scarcella and S. Krashen (Rowley, MA: Newbury House), 132–141.
- Gass, S. (1982). “From theory to practice,” in *On TESOL’81: Selected Paper from the Fifteenth Annual Conference of Teachers of English to Speakers of Other Languages*, eds M. Hynes and W. Rutherford (Washington, DC: TESOL), 129–139.
- Gibson, E., and Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Lang. Cogn. Process.* 28, 125–155. doi: 10.1080/01690965.2010.536656
- Gordon, P. C., Hendrick, R., Johnson, M., and Lee, Y. (2006). Similarity-based interference during language comprehension: evidence from eye tracking during reading. *J. Exp. Psychol.* 32, 1304–1321. doi: 10.1037/0278-7393.32.6.1304
- Hamilton, R. (1994). Is implicational generalization unidirectional and maximal? Evidence from relativization instruction in a second language. *Lang. Learn.* 44, 123–157. doi: 10.1111/j.1467-1770.1994.tb01451.x
- Hamilton, R. (1995). “The noun phrase accessibility hierarchy in SLA: determining the basis for its development effects,” in *Second Language Acquisition: Theory and Pedagogy*, eds R. E. Fred, M. Jean, R. W. Rita, H. Diane, and W. L. Peter (Mahwah, NJ: Lawrence Erlbaum Associates), 101–114.
- Hawkins, R. (1989). Do second language learners acquire restrictive relative clauses on the basis of relational or configurational information? The acquisition of French subject, direct object, and genitive restrictive clauses by second language learners. *Sec. Lang. Res.* 5, 155–188. doi: 10.1177/026765838900500204
- Holmes, V. M., and O’Regan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *J. Verb. Learn. Verb. Behav.* 20, 417–430.
- Hsiao, F., and Gibson, E. (2003). Processing relative clause in Chinese. *Cognition* 90, 3–27. doi: 10.1016/S0010-0277(03)00124-0
- Hsu, C.-C., and Chen, J.-Y. (2007). “A new look at the subject-object asymmetry: the effects of linear distance and structural distance on the processing of head-final relative clauses in Chinese,” in *Paper Presented at Interdisciplinary Approaches to Relative Clauses (REL07)*, (Cambridge, UK).
- Hyltenstam, K. (1984). “The use of typological markedness conditions as predictors in second language acquisition: the case of pronominal copies in relative clauses,” in *Second Languages: A Cross-Linguistic Perspective*, ed W. A. Roger (Rowley, MA: Newbury House), 39–58.
- Izumi, S. (2003). Processing difficulty in comprehension and production of relative clauses by learners of English as a second language. *Lang. Learn.* 53, 285–323. doi: 10.1111/1467-9922.00218
- Jeon, K. S., and Kim, H.-Y. (2007). Development of relativization in Korean as a foreign language: the noun phrase accessibility hierarchy in head-internal and head-external relative clauses. *Stud. Sec. Lang. Acquisition* 29, 253–276. doi: 10.1017/S0272263107070131
- Keenan, E. L., and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguist. Inquiry* 8, 63–99.
- Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., and Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: an eye-tracking study of pronominal relative clauses in Korean. *Language* 86, 546–582. doi: 10.1353/lan.2010.0006
- Lin, C.-J. C., and Bever, T. G. (2006a). “A universal account of relative clause processing in typologically diverse languages,” in *Abstract Presented at the 14<sup>th</sup> Annual Conference of the International Symposium on Chinese Language and Linguistics Joint Meeting*, (Taipei: Academia Sinica).
- Lin, C.-J. C., and Bever, T. G. (2006b). “Subject Preference in the processing of Relative Clauses in Chinese,” in *Proceedings of the 25th West Coast Conference on Formal Linguistics*, eds D. Baumer, D. Montero, and M. Scanlon (Somerville, MA: Cascadia Proceedings Project), 254–260.
- Lin, C.-J. C., and Bever, T. G. (2007). “Processing doubly-embedded head-final relative clauses,” in *Poster presented at Interdisciplinary Approaches to Relative Clauses (REL07)*, (Cambridge, UK: Cambridge University).
- Lin, C.-J. C., and Bever, T. G. (2011). “Garden path in the processing of head-final relative clauses,” in *Processing and Producing Head-Final Structures*, eds K. Hashimoto, J. Packard, and Y. Hirose (New York, NY: Springer), 277–297.
- Lin, Y. B., and Garnsey, S. M. (2011). “Animacy and the resolution of temporary ambiguity in relative clause comprehension in mandarin,” in *Studies in Theoretical Psycholinguistics: Vol. 38. Processing and Producing Head-Final Structures*, eds L. Frazier, T. Roeper, and K. Wexler (Springer), 241–275.
- Liversedge, S. P., Paterson, K. B., and Pickering, M. J. (1998). “Eye movements and measures of reading time,” in *Eye Guidance in Reading and Scene Perception*, ed G. Underwood (New York, NY: Elsevier), 55–75.
- Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., and Kentridge, R. W. (2004). Eye movements when reading disappearing text: is there a gap effect in reading? *Vision Res.* 44, 1013–1024. doi: 10.1016/j.visres.2003.12.002
- Miyamoto, E. T., and Nakamura, M. (2003). “Subject/object asymmetries in the processing of relative clauses in Japanese,” in *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, eds G. Garding and M. Tsujimura (San Diego, CA), 342–355.
- O’Grady, W., Lee, M., and Choo, M. (2003). A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Stud. Sec. Lang. Acquisition* 25, 433–448. doi: 10.1017/S0272263103000172

- O'Grady, W. (1997). *Syntactic Development*. Chicago, IL: University of Chicago Press.
- O'Grady, W. (2001). A linguistic approach to the study of language acquisition. *J. Pan Pac. Assoc. Appl. Linguist.* 5, 57–71.
- Ozeki, H., and Shirai, Y. (2007). "The consequences of variation in the acquisition of relative clauses: an analysis of longitudinal production data from five Japanese children," in *Diversity in Language: Perspectives and Implications*, eds Y. Matsumoto, D. Oshima, O. Robinson, and P. Sells (Stanford, CA: CSLI Publications), 243–270.
- Packard, J. L. (2008). Relative clause processing in L2 speakers of Mandarin and English. *J. Chin. Lang. Teach. Assoc.* 43, 107–146.
- Pavesi, M. (1986). Markedness, discursive modes, and relative clause formation in a formal and an informal context. *Stud. Sec. Lang. Acquisition* 8, 38–55. doi: 10.1017/S0272263100005829
- Rayner, K., and Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem. Cogn.* 14, 191–201.
- 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
- Schriefers, H. A., Friederici, and Kuhn, K. (1995). The processing of locally ambiguous relative clauses in German. *J. Mem. Lang.* 34, 499–520. doi: 10.1006/jmla.1995.1023
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., and Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *J. Mem. Lang.* 69, 104–120. doi: 10.1016/j.jml.2013.04.001
- Stites, M. C., Luke, S. G., and Christianson, K. (2013). The psychologist said quickly, "dialogue descriptions modulate reading speed!". *Mem. Cogn.* 41, 137–151. doi: 10.3758/s13421-012-0248-7
- Su, Y., Lee, S., and Chung, Y. (2007). Asyntactic thematic role assignment by Mandarin aphasics: a test of the trace-deletion hypothesis and the double dependency hypothesis. *Brain Lang.* 101, 1–18. doi: 10.1016/j.bandl.2006.12.001
- Sung, Y. T., Cha, J. H., Tu, J. Y., Wu, M. D., and Lin, W. C. (2015). Investigating the processing of relative clauses in Mandarin Chinese: evidence from eye-movement data. *J. Psychol. Res.* doi: 10.1007/s10936-015-9394-y. [Epub ahead of print].
- Tarallo, F., and Myhill, J. (1983). Interference and natural language processing in second language acquisition. *Lang. Learn.* 33, 55–76. doi: 10.1111/j.1467-1770.1983.tb00986.x
- Ueno, M., and Garnsey, S. M. (2008). An ERP study of the processing of subject and object relative clauses in Japanese. *Lang. Cogn. Process.* 23, 646–688. doi: 10.1080/01690960701653501
- Vasishth, S., Chen, Z., Li, Q., and Guo, G. (2013). Processing Chinese relative clauses: evidence for the subject-relative advantage. *PLoS ONE* 8:e77006. doi: 10.1371/journal.pone.0077006
- White, S. J. (2008). Eye movement control during reading: effects of word frequency and orthographic familiarity. *J. Exp. Psychol.* 34, 205–223. doi: 10.1037/0096-1523.34.1.205
- Wolfe-Quintero, K. (1992). Learnability and the acquisition of extraction in relative clauses and wh-questions. *Stud. Sec. Lang. Acquisition* 14, 39–70. doi: 10.1017/S0272263100010469
- Xu, Y. (2014). Evidence of the accessibility of hierarchy relative clauses in Chinese as a second language. *Lang. Linguist.* 15, 435–464. doi: 10.1177/1606822X14520666
- Yen, M.-H., Tsai, J.-L., Tzeng, O. J. L., and Hung, D. L. (2008). Eye movements and parafoveal word processing in reading Chinese. *Mem. Cogn.* 36, 1033–1045. doi: 10.3758/MC.36.5.1033
- Yip, V., and Matthews, S. (2007). Relative clauses in Cantonese-English bilingual children: typological challenges and processing motivations. *Stud. Sec. Lang. Acquisition* 29, 277–300. doi: 10.1017/S0272263107070143

**Conflict of Interest Statement:** 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 © 2016 Sung, Tu, Cha and Wu. 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) or licensor 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.



# Early literacy experiences constrain L1 and L2 reading procedures

Adeetee Bhide\*

Department of Psychology, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Rachel Helen Messer,  
Oklahoma State University, USA  
Ramesh Kalpa,  
Oklahoma State University, USA

### \*Correspondence:

Adeetee Bhide,  
Office 649, Learning Research and  
Development Center, 3939 O'Hara  
Street, Pittsburgh, PA, USA  
arb135@pitt.edu

### Specialty section:

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

**Received:** 02 April 2015

**Accepted:** 10 September 2015

**Published:** 02 October 2015

### Citation:

Bhide A (2015) Early literacy  
experiences constrain L1 and L2  
reading procedures.  
Front. Psychol. 6:1446.  
doi: 10.3389/fpsyg.2015.01446

Computational models of reading posit that there are two pathways to word recognition, using sublexical phonology or morphological/orthographic information. They further theorize that everyone uses both pathways to some extent, but the division of labor between the pathways can vary. This review argues that the first language one was taught to read, and the instructional method by which one was taught, can have profound and long-lasting effects on how one reads, not only in one's first language, but also in one's second language. Readers who first learn a transparent orthography rely more heavily on the sublexical phonology pathway, and this seems relatively impervious to instruction. Readers who first learn a more opaque orthography rely more on morphological/orthographic information, but the degree to which they do so can be modulated by instructional method. Finally, readers who first learned to read a highly opaque morphosyllabic orthography use less sublexical phonology while reading in their second language than do other second language learners and this effect may be heightened if they were not also exposed to an orthography that codes for phonological units during early literacy acquisition. These effects of early literacy experiences on reading procedure are persistent despite increases in reading ability.

**Keywords:** orthographic depth hypothesis, whole word, phonics, ESL, word recognition

## Introduction

Models of word reading have broadly identified two pathways to word recognition: first accessing pronunciation or first accessing meaning. Everyone uses both pathways to some extent while reading, but the division of labor between them (i.e., *reading procedure*) can vary depending on the word type, context (Besner and Smith, 1992), the early literacy experiences of the individual, etc. In this review, I demonstrate that early literacy experiences have a profound effect on reading procedure. I begin by briefly reviewing differences between orthographies and models of word recognition. I then demonstrate the first language (L1) persistency effect, that effects of L1 orthographic transparency and instructional method are measureable, not only in beginning readers, but also in highly skilled adult readers. Finally, I demonstrate the L2 persistency effect, that that early literacy experiences are able to exert an effect even while one is learning to read a second language, and that these effects remain with increasing L2 proficiency. Although much work has been done on the effect of instructional method on reading procedure (Connelly et al., 2009) and on the effect of L1 transparency on both L1 (Katz and Frost, 1992) and L2 reading procedure, this review is unique in that it brings together these three lines of research into one integrated framework.



## Transparency

Writing systems are defined by the phonological grain size that each graph represents (see **Table 1**). Alphabetic orthographies, such as English, Serbo-Croatian, Korean<sup>1</sup>, Russian, German, and French, have graphs that code for phonemes. Alphasyllabic orthographies, such as Hindi, Marathi, and Thai, have graphs that code for syllables but subcomponents of the graphs code for phonemes. Abjads (e.g., Hebrew, Persian, and Arabic) are similar to alphasyllabaries but some<sup>2</sup> vowel subcomponents are typically excluded in text. Syllabic orthographies, such as Japanese hiragana and katakana (collectively called kana), have graphs that code for syllables. Finally, morphosyllabic orthographies, such as Chinese and Japanese kanji<sup>3</sup>, have graphs that code for morphemes.

<sup>1</sup>Unlike other alphabets, Korean uses a non-linear graph arrangement. Also, note that although Korean Hangul is alphabetic, Korean occasionally uses some Chinese characters as well (Taylor and Taylor, 1995).

<sup>2</sup>In unvowelized Hebrew, almost all vowels are orthographically unexpressed. Two graphs can, in certain contexts, express vowels, however, these graphs also express consonants (Frost, 2012). In unvowelized Arabic, long vowels are orthographically expressed but short vowels are not (Fender, 2008a).

<sup>3</sup>Note that Japanese kanji is not, strictly speaking, morphosyllabic. Most kanji characters have two readings, on and kun. The on-reading is the Chinese pronunciation whereas the kun-reading is the Japanese pronunciation. Although the on-reading is monosyllabic, the kun-reading may be multisyllabic. Furthermore, because kanji graphs have multiple pronunciations, there is low character-phonological unit consistency (unlike Chinese).

Phonological transparency refers to how systematically a given graph maps onto a given phonological unit and vice versa (see **Table 1**). Although languages with the same writing system can vary in terms of transparency, transparency is not independent of writing system. Alphabets range from highly transparent (e.g., Serbo-Croatian) to moderately opaque (e.g., English). Note that grapheme-phoneme and phoneme-grapheme consistencies are not necessarily equivalent; for example, French and German have higher grapheme-phoneme consistency than phoneme-grapheme consistency (Deacon et al., in press; Landerl, in press).

Alphasyllabaries have high graphemic subcomponent-phoneme consistency; graphemic subcomponents typically map onto only one phoneme. There is less phoneme-graphemic subcomponent consistency because the schwa vowel is inconsistently represented (Bhide et al., 2013). Furthermore, the visual form of a graphemic subcomponent may change depending on other subcomponents in the graph. Syllabaries tend to be highly transparent. The vowelized versions of abjads are highly transparent. In the unvowelized form, the grapheme-phoneme correspondences are highly transparent<sup>4</sup>. However, some vowel phonemes are not orthographically represented, so the phoneme-grapheme correspondences are

<sup>4</sup>In Hebrew, some diacritical marks are used to distinguish consonants. These diacritical marks are excluded in the unvowelized form, making the consonants slightly ambiguous (Frost and Bentin, 1992).

**TABLE 1 | The trade-off between phonology and morphology for the word “money” in various orthographies.**

Orthography Writing system “money”	Graph-phonological unit consistency	Phonological unit-graph consistency	Trade-off between phonology and morphology (see Perfetti and Harris, 2013)
Japanese hiragana Syllabary おかね /okane/	Most kana graphs have unique pronunciations, but there are exceptions (see Akita and Hatano, 1999)	Highly consistent	Precedence is given to preserving phonology
Spanish Alphabet moneda /moneda/	Fairly consistent but there are exceptions (see Defior and Serrano, in press)	Fairly consistent but there are exceptions (see Defior and Serrano, in press)	Precedence is given to preserving phonology; e.g., the “d” changes to a “t” in moneda → monetario (monetary)
Korean Alphabet 돈 /don/	Highly consistent	Highly consistent	Phonological transparency has been sacrificed in a few words to make morphological relationships more apparent (see Perfetti and Harris, 2013)
Marathi Alphasyllabary पैसा /paisa/	Highly consistent	Fairly consistent, but the visual forms of graphs change depending on orthographic context. The /a/ is not always orthographically represented.	Precedence is given to preserving phonology; e.g., the स /s/ changes to a श /ʃ/ in पैसा → पैशासंबंधी (monetary)
English Alphabet money /mʌni/	Inconsistent, especially for vowels	Inconsistent	Phonology is sometimes sacrificed for morphology; e.g., money and monetary both use the letter “o,” although its pronunciation changes
Hebrew Abjad קֶסֶף /kesef/	Highly consistent	Vowels are not orthographically represented. The visual forms of graphs can change based on word position.	Exclusion of vowels makes morphological relationships more apparent; e.g., money (which also means silver) shares a consonantal root structure with gray-haired (כֶּסֶף שֵׁנִי /kisuf se'ni/) and silver-plated (הַכֶּסֶף הַזָּהָבִי /haskesef hazahavi/)
Chinese Morphosyllabary 钱 qian2	Typically consistent at multi-character level; inconsistent at the character level; e.g., the phonetic radical in “money” can be found in characters that have various pronunciations such as jian1, jian4, pan4, can2, and zhan4	Extremely inconsistent: many characters represent the same syllable; e.g., 前, 钳, and 乾 are all pronounced qian2	Phonological transparency is sacrificed for semantic information; e.g., the character for “money” has the semantic radical “gold.” Inclusion of semantic information distinguishes between homophones.

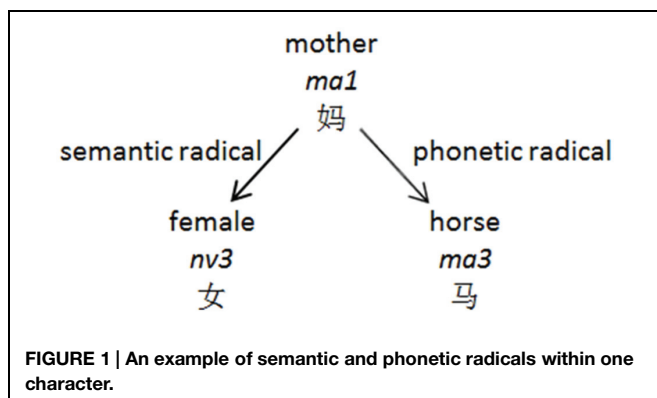
highly opaque. Furthermore, the visual forms of graphs can change based on word position (Saiegh-haddad, in press).

In Chinese, a morphosyllabary, 80–90% of characters are made up of two components, a phonetic and semantic radical, which provide a cue to the character's pronunciation and meaning, respectively (Kang, 1993 as cited in Li and McBride-Chang, 2014) (see **Figure 1**). However, a given phonetic radical can cue for multiple pronunciations and the same syllable can be cued by multiple phonetic radicals. Therefore, at the character level, morphosyllabaries are so highly opaque that they are considered an outlier orthography. However, multi-character, multi-syllabic words are highly transparent; each character typically has one pronunciation and represents one syllable in the word.

Opacity can stem from multiple sources: some phonemes may not be represented in the text (in the case of Hebrew) or multiple graphemes may correspond to the same syllable (in the case of Chinese). Readers of Hebrew must rely heavily on context to disambiguate the many homographs whereas readers of Chinese must have highly specified orthographic knowledge to distinguish between homophonic characters.

Opacity does not necessarily make an orthography more difficult to read. There is a trade-off between accurately representing phonological and morphological information and opacity often results from the inclusion of more morphological information at the expense of phonology [Perfetti and Harris (2013); see **Table 1**]. For example, the vowel digraph “ea” is pronounced differently in the words “heal” and “health,” making the words opaque. However, this phonological ambiguity makes the semantic similarity more apparent. The opacity also allows English to distinguish between homophones such as “heal” and “heel”. In abjads, such as Arabic, sequences of three consonants are used to form roots, which can be combined with a variety of vowels to form families of semantically related words. For example, the root k-t-b is found in the words ketaab (book), kataba (he wrote), and maktaba (library).

These 3-letter roots are more apparent in the unvowelized versions (Ryan and Meara, 1991), another example of how phonological opacity allows for more morphological clarity. In morphosyllabaries, the characters are highly opaque, but provide a great deal of semantic information via the semantic radical. The character system also allows for the disambiguation of the many homophones in Chinese and Japanese.



## Models of Word Reading

This basic division of labor can be simulated by computational models of word reading. The DRC model (Coltheart et al., 2001) models word naming and posits that there are two possible pathways: (1) the sublexical phonological route that uses letter-sound correspondence rules to sound out words and (2) a lexical route that maps the orthographic form onto a stored whole-word phonological representation. The pathways share a common speech output (phoneme) system called the response buffer<sup>5</sup>. If both pathways activate the correct phonemic sequence, the response is faster than if the pathways are in competition.

A connectionist alternative to the dual-route model is the triangle model (Seidenberg and McClelland, 1989). The triangle model was initially developed to model word naming, and also claimed that there are two pathways. The first pathway uses statistical regularities of sublexical constituents to activate phonological features, whereas the second maps the orthographic form onto its semantic features, and from there accesses the phonological features. Later, the triangle model was applied to examine how word meaning is computed (Harm and Seidenberg, 2004). Although there are differences between the models, both models posit that words can be read using either sublexical phonological information or morphological information derived from analyzing morphemes as orthographic wholes. This review examines the division of labor between the two pathways.

## Assessing the Division of Labor

Although many different experimental manipulations can be used to examine reading procedure, there are seven *signature manipulations* that have been used in many studies. These *signature manipulations*, as well as other experimental manipulations/paradigms, are used throughout the review to examine reading procedure.

If the pattern of results outlined below are found in studies, it would suggest that people are heavily relying on the sublexical phonological pathway:

- (1) *Lexicality: if words and non-words are named with the same speed and accuracy.* Words are read faster than non-words because words benefit from the use of both routes. If words do not show an advantage over non-words, it suggests that the morphological pathway is being underutilized during word reading (Frost et al., 1987).
- (2) *If large effects of length are seen.* Because the DRC model claims that grapheme-phoneme correspondences are computed in a left to right fashion<sup>6</sup> (Coltheart et al., 2001), longer words should take more time to decode via the sublexical phonological pathway.
- (3) *Regularity: if readers are very slow to name exception words or make many regularization errors.* According to the DRC model, if the sublexical phonological pathway is heavily weighted while reading exception words, it will slow

<sup>5</sup>Note that there are other dual route models which posit that all stimuli are processed by either the lexical or sublexical pathway and the pathways do not interact (see Norris and Brown, 1985) but there is considerable evidence against those models (Coltheart et al., 2001).

<sup>6</sup>In linear alphabets that are read from the left to the right.

down the lexical pathway or produce regularization errors (Coltheart et al., 2001). According to the triangle model, the sublexical phonological pathway is slower and less accurate at naming exception words<sup>7</sup> (Seidenberg and McClelland, 1989).

- (4) *If homophone/pseudohomophone effects are seen.* According to the triangle model, the phonological pathway cannot distinguish between homophones, pseudohomophones, and exemplars (Harm and Seidenberg, 2004). Note that studies employing homophones and pseudohomophones use various paradigms (e.g., lexical decision, backward masking, text reading, definition selection, naming, semantic judgment, spelling recognition, etc.), but share an underlying theoretical logic.

In contrast, if the pattern of results outlined below is found in studies, it would suggest that people are largely using morphological/orthographic information:

- (1) *If large frequency effects are seen in lexical decisions or naming.* According to the DRC model, entries in the orthographic lexicon are frequency-sensitive (Coltheart et al., 2001). Simulations with the triangle model demonstrated that the semantic pathway is more sensitive to frequency than the phonological pathway is (Harm and Seidenberg, 2004).
- (2) *If large effects of imageability or semantic priming are seen during naming tasks.* Imageability is a semantic variable and semantic priming preactivates relevant semantic features (Katz and Frost, 1992; Harm and Seidenberg, 2004).

It is important to note that reading procedure is different than reading ability. Different reading procedures entail differing emphases on the two reading pathways. In contrast, reading ability refers to overall differences in reading accuracy and/or speed. This review focuses on reading procedure and hence does not examine whether particular orthographies or teaching methods are associated with higher reading ability. Rather, it examines whether particular orthographies or teaching methods are associated with different reading procedures, and to get a more pure measure of reading procedure, it primarily focuses on studies in which the samples are matched for overall reading ability. For example, differences in non-word reading accuracy are only meaningful when the groups are matched for word reading accuracy. Other measures of reading procedure, such as error type analysis, can give us some information about reading procedure even when reading ability is not matched, but the results should always be interpreted with caution.

## Reading in the L1

### Transparency of the L1 Orthography

Previous studies suggest that people whose L1s have transparent orthographies use more sublexical phonology while reading than do people whose L1s have opaque orthographies, and that this is true for both beginning and skilled readers. The studies

have compared English to more transparent orthographies (e.g., Serbo-Croatian, German, Albanian, Greek, Japanese hiragana, Welsh) or to more opaque orthographies (e.g., Hebrew, Chinese, Japanese kanji). They used signature manipulations, such as lexicality, homophonic/pseudohomophonic status, length, frequency, and semantic priming, to examine reading procedure.

### Beginning Readers

A comparison between 7 and 9 years old children learning German and English, matched for word reading skill, found that English-speaking children struggled more with reading non-words aloud than did German-speaking children. Because the children were also matched for years of schooling, this introduced age as a possible confounding factor (Wimmer and Goswami, 1994). However, the results are robust as another study was able to find the same effect using a different paradigm and without the age confound; Goswami et al. (2001) found that German-speaking children showed more pseudohomophonic interference and a greater length effect on a lexical decision task than did English-speaking children (matched for reading and spelling).

Other studies using larger age ranges (5–15 years old) and more diverse language groups (Welsh, Albanian, Greek, Japanese, and English) have corroborated the general pattern of results described above. These studies did not match the participants as well as the Goswami studies did, making the results of each individual study less convincing. However, taken as a whole, the research does suggest an effect of orthographic transparency on reading procedure. The studies found a stronger relationship between word length and naming time in readers of transparent as compared to opaque orthographies (Ellis and Hooper, 2001; Ellis et al., 2004). Furthermore, error types vary by orthography. Children reading transparent orthographies were more likely to make mispronunciations that result in non-words (e.g., saying “polical” for “political”), which is consistent with a phonological assembly strategy. In contrast, children reading opaque orthographies were more likely to make whole word substitution errors (e.g., saying “computer” for “complete”), which is consistent with a reading strategy that attempts to map the whole visual form onto a lexical entry (Ellis and Hooper, 2001; Spencer and Hanley, 2003; Ellis et al., 2004). However, as mentioned above, there are some problems with matching participants. Primarily, the participants in these studies were not matched for reading ability; children reading transparent orthographies had higher word reading scores than those reading more opaque orthographies. Hence, it is possible that these differences in reading procedures would diminish when using ability-matched samples. Furthermore, although the participants in the Ellis and Hooper (2001) and Spencer and Hanley (2003) studies comparing English and Welsh speaking students were relatively well matched demographically<sup>8</sup>, the participants in the

<sup>7</sup>Especially low frequency exception words.

<sup>8</sup>Ellis and Hooper (2001) matching criteria: the students all lived in the Wrexham area of Wales, their schools were all in similar in terms of catchment area, classroom size and teaching method, the students were all in Year 2, the students were matched in terms of math ability. Spencer and Hanley (2003) matching criteria: the students were in local education authority schools in Denbighshire, North Wales, the students were all in Year 2, the schools used phonics to teach literacy.

Ellis et al. (2004) study comparing English, Japanese, Greek, and Albanian speaking students were not well matched in terms of recruiting and testing procedures as well as cognitive abilities.

### Skilled Readers

In addition to the effects of orthographic transparency on beginning readers, as detailed above, effects of orthographic transparency have also been found among skilled readers, providing evidence for the L1 persistency effect. Readers of more transparent orthographies tend to read non-words more accurately and are less affected by lexicality on their naming reaction times (RTs) than are readers of more opaque orthographies. Furthermore, semantic primes during naming tasks benefit readers of more opaque orthographies to a greater degree than readers of transparent orthographies (Katz and Feldman, 1983; Frost et al., 1987).

Another way of measuring reading procedure is by forcing strategy changes. Frost et al. (1987) had undergraduate students perform a naming task with priming. In Hebrew, naming was significantly slowed when the prime was a non-word. A smaller effect in the same direction was found for English, but no effect was found for Serbo-Croatian. In another experiment, participants were asked to name words and non-words as quickly as possible. The authors varied the proportion of non-words within the lists. Hebrew speakers were significantly less accurate when there was a high proportion of non-words in the list, a trend in the same direction was seen with English speakers, whereas no effect was seen for Serbo-Croatian speakers. Non-word primes and high proportions of non-words force people to use the sublexical phonological pathway. Hebrew readers are accustomed to using the morphological pathway, and the forced strategy change slows them down. In contrast, Serbo-Croatian readers typically use phonological information, so no strategy change is needed.

The Frost et al. (1987) study demonstrated that readers of shallow orthographies use phonology to identify a word, whereas readers of opaque orthographies access phonology once a word is identified. Studies testing English and Chinese participants using a backward masking paradigm came to the same conclusion (Perfetti and Bell, 1991; Perfetti and Zhang, 1991; Perfetti et al., 1992). In the paradigm, a word target is briefly shown, followed by a brief prime, and then a pattern mask. The English-speaking participants were more accurate at identifying the word target when the prime was a pseudohomophone of the word than an orthographic control (e.g., more accurate at identifying “rate” in the condition rate-RAIT-XXXX than in the condition rate-RALT-XXXX). In contrast, priming the target character with a homophonic character did not improve the accuracy of the Chinese participants. The interpretation of these findings is that, in English, the partial products of target identification include pre-lexical phonology, so phonological primes are able to reinstate the partial products. In Chinese, there is no pre-lexical activation of phonology, so there is no effect of the phonological prime.

One possible problem with comparing the aforementioned English and Chinese studies is that the English speaking participants were on average younger (all undergraduates) than

the Chinese speaking participants (mostly graduate students). Furthermore, the English tasks used pseudohomophone primes whereas the Chinese tasks used homophone primes. Another English-Chinese comparison used tighter age controls (all undergraduates) and homophones for both languages and also found that English speakers use phonology during word identification whereas Chinese speakers access phonology after word identification. However, in contrast to the studies cited about, this study looked at participants reading texts, rather than individual words. Feng et al. (2001) measured eye movements while participants read texts in which some words were either retained or replaced either by their homophone or an orthographic control (e.g., “creek” was replaced either by “creak” or “creed”). For the English texts, distributional analyses revealed that, for first fixations less than 200 ms, homophones were indistinguishable from the targets, both of which differed significantly from the orthographic controls. This was especially true if the homophone and the original word had a high degree of orthographic overlap and if the target was highly predictable from context. In contrast, when reading Chinese texts, participants had longer first fixation durations for all orthographic mismatches. This difference suggests that, while reading English, participants use phonology, orthography, and context to identify words. In contrast, while reading Chinese, people mainly use orthography.

### Instructional Method

In addition to orthographic transparency, instructional method can also influence reading procedure. There are numerous instructional philosophies for teaching literacy. Broadly, the instructional methods cluster into two groups; phonology-based methods (such as phonics) and semantics-based methods (such as the whole word method). Phonology-based methods all focus on sounding out words, although the grain size of focus can vary (see Brown and Deavers, 1999; Ziegler and Goswami, 2006; Asfaha et al., 2009; Nag, 2011; Kyle et al., 2013). In contrast, semantics-based methods focus on recognizing whole words and deriving meaning from text. The goal of this section is to compare phonology and semantics based methods to determine if they are associated with different reading procedures. I chose not to compare phonological methods that focus on different grain sizes on the assumption that they all foster a reading procedure that is more dependent on the phonological pathway. However, there is some preliminary evidence from an artificial orthography study that phonological methods focusing on larger grain sizes may foster reading procedures that rely more heavily on the morphological pathway as compared to phonological methods focusing on smaller grain sizes (Hirshorn et al., 2015). There is not enough current research to examine the effect of teaching phonology at different grain sizes on reading procedure in the present review, but it is an interesting area of further inquiry.

Almost all of the studies focusing on reading procedure as an outcome have compared the phonics and whole word instructional methods, so I discuss those in detail. In the phonics method, letter-sound correspondences are introduced in a systematic manner. The whole word method (or book experience in New Zealand), emphasizes communication and comprehension. Words are often memorized from texts

proposed by students and letter-sound correspondences are not taught systematically. The majority of studies has been done with English, so I focus on English in this section and then examine how generalizable the conclusions are to other languages further on.

Studies were included in this section only if they compared people who learned English via the phonics and whole word methods and used signature manipulations to examine reading procedure. The signature manipulations used include lexicality, regularity, length, frequency, pseudohomophonic/homophonic status, and imageability manipulations. Many of these studies (see Connelly et al., 2009 for review) compared students from Scotland, whose curriculum stresses phonics instruction, and New Zealand, whose curriculum stresses book experience. Although these studies use a cross-national sample, Scotland and New Zealand have close cultural ties and their educational systems share a common history and educational culture. Therefore, differences found between the samples are likely due to instructional differences. In the systematic phonics approach in Scotland, individual letter-sound correspondences are explicitly taught and students are encouraged to use sequences of such correspondences to sound out unfamiliar words. In New Zealand, almost all of the literacy instruction centers on story texts. Teachers help students understand the meaning of the story, and teach them to recognize words using context cues, initial letters/letter clusters, and analogies to other words. However, they are never taught to sound out successive letters.

### Beginning Readers

Studies with 6–8 years old children have shown that students taught with a phonics focus typically outperform students taught with a book experience focus on *non-word* reading tasks, even while controlling for word reading ability and various demographic factors such as age, years of schooling, socioeconomic status (SES), word recognition, aural vocabulary, spelling ability, and short term memory (Thompson and Johnston, 2000; Connelly et al., 2001; Thompson et al., 2008). In addition to finding differences in non-word reading accuracy, Connelly et al. (2001) also found differences for *word* reading accuracy. Specifically, children taught with a phonics focus named regular words more accurately, but children taught with a book experience focus named exception words more accurately. The authors also compared the children on their ability to name highly familiar words which they encounter daily (e.g., “the,” “he”). The children with phonics instruction read the words significantly slower than the children with the book experience instruction, suggesting that they were less successful at incorporating these words into their sight vocabularies. Connelly (unpublished thesis, as cited in Connelly et al., 2009) found a length effect in naming words with 5–6 years old children from Scotland but not from New Zealand (the children were matched for overall reading accuracy), again suggesting that children with explicit phonics instructions use more phonological recoding while reading.

Johnston and Thompson (1989) compared the performance of 7–8 years old children from Scotland and New Zealand

(matched for age and reading ability, differences in vocabulary size were statistically controlled for) on a lexical decision task containing words, pseudohomophones, and non-words. The New Zealand children were equally accurate at rejecting the non-words and pseudohomophones, whereas the Scottish children were more accurate at rejecting the non-words than the pseudohomophones.

Finally, Connelly et al. (2001) found that children taught with a phonics focus are more likely to attempt to sound out unfamiliar words than are children taught with a book experience focus. The 6–7 years old children from Scotland were more likely to produce non-word errors or contextually appropriate errors that retained the pronunciation of at least two of the letters in the original word than were the children from New Zealand.

All of the English studies cited above used a cross-national sample to examine the effect of instructional method, so it is possible that there were some socio-cultural confounds. Two studies were able to minimize confounding factors by comparing instructional method within the same country. Foorman et al. (1991) compared first graders in the U. S. receiving either more or less letter-sound instruction in their school curriculums. The students were matched for reading ability, vocabulary, SES, and ethnic diversity. However, there were a couple of notable differences: the students receiving less letter-sound instruction were drawn from public schools and were on average 2 months older than the students receiving more letter-sound instruction, who were drawn from parochial schools. The authors found that the students receiving more letter-sound instruction showed larger regularity effects while reading aloud.

Whereas Foorman et al. (1991) compared students in the U. S., Landerl (2000) compared students living in England. The study was designed as a follow-up to the Wimmer and Goswami (1994) study cited previously. In the original study, children learning German were compared to children learning English to look for an effect of orthographic transparency. However, the children learning English were taught using a blend of the whole-word and phonics methods, whereas the children learning German were primarily taught using the phonics method. In the follow-up study, two groups of English-speaking children were used, one that received a mix of whole word and phonics instruction and one that received primarily phonics based instruction. They were compared to German-speaking children who received phonics based instruction. The three groups were equivalent in terms of their word reading. However, for non-word reading, the German-speaking children performed the best, followed by the English-speaking children receiving the phonics based instruction, whereas the English-speaking children receiving the mixed instruction performed the worst.

### Skilled Readers

The studies cited above demonstrate that instructional method affects the reading procedure of children just beginning to read (5–8 years old), even while controlling for various confounding factors. However, the question that remains is, are these differences in reading procedure stable over time (i.e., can we demonstrate the L1 persistency effect)? It is possible that as reading skill increases, more words are added to sight vocabulary,

reducing the need for sublexical phonology. Furthermore, as reading skill increases, all adults, despite how they were taught to read, may settle on the same “optimal” reading procedure. The research (reviewed below) suggests that this is not the case. The ability of specific tasks to detect differences in reading procedure fluctuates with age, but some effect of instructional method on reading procedure is measurable in both adolescence and adulthood.

Johnston et al. (1995) examined whether differences in reading procedure remain constant across adolescence. The participants included both 8 and 11 year-old children from Scotland and New Zealand who were matched on reading ability, vocabulary, chronological age, and ethnicity. They were tested on a lexical decision task that included words, pseudohomophones, and non-words. Unlike the 8 year-olds taught with the book experience method, the 8 year-old children taught with the phonics instruction were less accurate at correctly rejecting the pseudohomophones than the non-words. Both groups of 11-year olds were equally accurate at rejecting pseudohomophones and non-words. Therefore, for this task, the effect of instructional method diminishes with reading experience. The participants also completed a pseudohomophone sentence evaluation task, where they have to judge whether or not a sentence is orthographically correct. The incorrect sentences had one word replaced by either a pseudohomophone (e.g., Can you poast this letter?) or by a control non-word (e.g., She has loast her bag.). Both groups of 8 year-olds were less accurate at evaluating the sentences with a pseudohomophone than the sentences with a control non-word. The 11 year-olds taught with the phonics method (but not the 11 year-olds taught with the book experience method) were also less accurate at evaluating the sentences with a pseudohomophone. Therefore, for this task, the effect of instructional method becomes more apparent with age. Reading instruction also had an effect on the participants’ accuracy in a word-meaning task, where participants had to choose the correct definition for presented words. Some of the stimuli were homophonic (e.g., son) and their definition (e.g., child) and the definition of their homophone (e.g., light) were among the choices. Children taught with the phonics method made more errors when choosing the correct definition for the homophonic stimuli than did the children taught using the book experience method. Therefore, on this task, the effects of teaching methodology seem relatively stable during early adolescence.

Thompson et al. (2009) were able to find effects of instructional method even among skilled adults. They compared university students (matched for age and vocabulary) who had learned to read in Scotland and New Zealand on non-word reading. The non-words included irregular, body-consistent stimuli (e.g., thild) where two responses are legitimate, the regular response that is inconsistent with the bodies of all real words and the pronunciation that is consistent with the bodies of words such as “mild”. Although the two groups were equally accurate, the Scottish adults were more likely to give regular responses and less likely to give irregular responses than were the New Zealand adults. These results are easiest to explain using the DRC model. The sublexical route uses grapheme-phoneme

correspondence rules to sound out a word, without taking into account the greater orthographic context. So, for “thild,” the sublexical route would read it as /θɪld/. In contrast, the lexical route would activate orthographically similar words, such as “child,” “mild,” “thick,” and “third”. These lexical representations would activate their phonological representations which would then activate the phonemes within them. Therefore, /aɪld/ and /θ/ would be highly active, producing /θaɪld/. Therefore, greater dependence on the sublexical phonological route produces the regular response, whereas greater dependence on the lexical route produces the irregular response. These data are harder to explain with the triangle model because the sublexical route uses statistical regularities that are sensitive to orthographic context. Therefore, the triangle model is more likely to produce irregular responses.

Another result from the same study allows us to more easily interpret reading procedure using both computational models. The participants were asked to name words that varied in terms of both frequency and imageability. The Scottish adults were more likely to make regularization errors while reading low frequency, low imageability words than were the New Zealand adults. These results suggest that, for Scottish adults, the word types that engage the lexical/semantic pathways to the smallest degree were unable to elicit enough support from the lexical/semantic pathways to avoid regularization errors.

### Summary and Extension to Other Languages

Overall, the research suggests that readers of transparent orthographies (at all levels of reading skill) heavily weight the sublexical pathway while reading. In contrast, readers of more opaque orthographies rely more on morphological/orthographic information. The research on this topic has compared English to more transparent alphabets and syllabaries (German, Serbo-Croatian, Welsh, Albanian, Greek, and Japanese hiragana) and to more opaque abjads and morphosyllabaries (Hebrew, Chinese, and Japanese kanji), and found that the general conclusion held for all of those orthographies. There appear to be no studies using alphasyllabaries, but we can predict that readers of alphasyllabaries heavily weight the sublexical phonology pathway because they are transparent.

Studies of English have suggested that, in addition to transparency, instructional method can also influence reading procedure. Students taught with a phonics-based focus more heavily weight the sublexical phonology pathway, whereas students taught with a book-experience or whole-word focus weight the morphological pathway more heavily. The question that remains is, is this conclusion generalizable to other languages, or is it English specific? The evidence reviewed below suggests that the conclusion is applicable to other opaque orthographies such as Chinese, but not to more transparent orthographies such as French.

Leybaert and Content (1995) compared two French-speaking schools in Belgium (matched for SES) that used different teaching methods and examined which pathway (sublexical phonological or morphological) was more heavily weighted in their students. The authors compared age and ability-matched samples in separate analyses to control for reading experience and reading

skill, respectively. Four different reading tests were administered in which different variables were manipulated to tease apart the weightings given to different pathways. The first test manipulated word regularity<sup>9</sup>, the second test manipulated both complexity of the grapho-phonological correspondences<sup>10</sup> and lexicality, the third test varied the frequency of words and the length of words and pseudowords, and the final test contained words, and homophonic and non-homophonic pseudowords. For all the reading tests, participants had to read the items aloud as quickly as possible.

When comparing the reading ability-matched groups, the effect of teaching methodology was only visible on one of these reading tests; on the test that varied both grapho-phonological complexity and lexicality, there was an interaction between teaching method and lexicality in that the students receiving whole word instruction were slightly more accurate on the words and less accurate on the non-words than the students receiving phonics instruction. When comparing age-matched groups, the fourth and sixth graders receiving phonics instruction showed a larger regularity effect on both RT and accuracy than those receiving whole word instruction<sup>11</sup>. Overall, the Leybaert and Content (1995) study showed a minimal effect of teaching methodology in developing readers; out of four reading tests intended to measure the weightings given to the two pathways, only one showed a significant effect of teaching methodology using age-matched groups and another using ability-matched groups. And, on the test that found effects using ability-matched groups, only a methodology x lexicality interaction was found, not a methodology x lexicality x graphophonological complexity effect. A three-way interaction was predicted because graphophonological complexity affects decoding, so its effect should be larger for non-words than for words only if participants heavily use the morphological pathway when reading words.

Instruction had relatively little influence in French, which has a transparent alphabet, but it had a significant effect for English, which has a more opaque alphabet. The results for Chinese, a highly opaque morphosyllabary, echo those of English: instruction is able to exert an effect. As stated previously, most Chinese characters contain two components, a phonetic and semantic radical (see **Figure 1**). Unlike the DRC model, which cannot handle Chinese due to its lack of grapheme-phoneme correspondence rules, the triangle model can be trained such that the sublexical pathway can learn the statistical regularities among the phonetic radicals, whereas the semantic pathway can learn the meaning of the whole character, with the help of the semantic radicals (Yang et al., 2006, 2008). Different instructional methods have been shown to favor one reading method over another.

<sup>9</sup>Regular words conform in pronunciation to the most frequent correspondences between speech sounds and letters/groups of letters. Irregular words include spelling patterns that deviate from their most common pronunciation.

<sup>10</sup>Items with low graphophonological complexity have letters that only map onto one phoneme, independent of orthographic context. Complex items contain phonemes that can be represented by more than one letter or letters that can represent more than one phoneme depending on orthographic context.

<sup>11</sup>Although some regularity effects were also seen with the second graders, these are difficult to interpret due to their low accuracy.

**TABLE 2 | Several words written in traditional Chinese characters, as well as the alphabetic orthographies pinyin (used in Mainland China) and Zhu-Yin-Fu-Hao (used in Taiwan).**

Character	Meaning	Pinyin	Zhu-Yin-Fu-Hao
媽	Mother	mā	ㄇㄚ
抹	Dust (v)	mā	ㄇㄚ
膜	Film (thin covering)	mó	ㄇㄛˊ
辣	Spicy	là	ㄌㄚˋ

*Characters are able to distinguish between the homophones “mother” and “dust,” whereas the alphabetic orthographies are not.*

In Hong Kong, Chinese is taught using the “whole word method.” Characters are taught through rote copying and in the context of texts. Children are encouraged to rapidly identify whole characters. In contrast, in Taiwan, teachers are more likely to draw attention to the phonetic radicals and to teach characters in phonologically related sets. Furthermore, an alphabetic system called Zhu-Yin-Fu-Hao is used to phonologically transcribe characters (**Table 2**) (Scholfield and Chwo, 2005).

Scholfield and Chwo (2005) studied sixth grade students from Taiwan and Hong Kong whose schools were of similar sizes and served socioeconomically comparable populations. They presented two characters to the students and asked them to make a meaning similarity judgment. Some of the foils were phonologically similar, whereas others were graphically similar. The students from Taiwan were slowed to a greater extent by the phonologically similar foils, whereas the students from Hong Kong were slowed to a greater extent by the graphically similar foils. This finding suggests that the instructional method in Taiwan leads to a greater dependence on phonological recoding during character recognition.

The research so far suggests that instructional method has a greater effect on languages with opaque orthographies, such as Chinese and English, than on languages with more transparent orthographies, such as French. More studies are needed to confirm this general conclusion because, to the best of my knowledge, only one study has examined the effect of instruction on a transparent orthography. It is also possible that instruction may influence reading procedure when beginning to read a transparent orthography, but not after a critical level of fluency has been reached; some preliminary data with learning to read a transparent artificial orthography show an effect of instruction on reading procedure (Taylor et al., 2015). A review of the literature did not reveal any studies that have looked at the effect of instruction on learning to read an alphasyllabary, syllabary, or abjad. We can predict that, for transparent alphasyllabaries and syllabaries, teaching method should have little effect on the weightings given to the two pathways. In fact, because syllables are more salient than phonemes in spoken language (Ziegler and Goswami, 2005), instructional method may have less of an effect in transparent syllabaries than in transparent alphabets. Because abjads are opaque, I predict that teaching methodology should be able to exert an effect on reading procedure, but more research is needed to confirm this hypothesis. It is possible that other factors besides transparency [such as the size of the graphemic set,

see Nag (2011)] could moderate the influence that instructional method has on reading procedure.

## Reading in the L2

The language one was first taught to read, and how he/she was taught, can have powerful effects on reading procedure well into adulthood. Furthermore, early literacy experiences can even affect one's approach to reading in a foreign language. L1–L2 transfer effects have been broadly studied in the literature and evidence of transfer has been found at all the levels of the language system (MacWhinney, 2001). Although these broad transfer effects are outside the scope of this paper, I demonstrate that the transparency of one's L1 orthography can have effects even while reading in a second language and these effects remain even with increasing L2 proficiency. Furthermore, the instructional method of L1 literacy may also affect reading procedure. Similar to the L1 research, the L2 research has largely focused on learning English. Therefore, I begin with the effect of L1 orthographic transparency and instructional method on learning to read English and then examine how generalizable the conclusions are to other languages later on.

### Transparency of the L1 Orthography

The studies included in this section have all compared L1 readers of morphosyllabaries to L1 readers of more transparent orthographies or to native English speakers. They have found that when L1 readers of morphosyllabaries learn a more transparent orthography, they use less sublexical phonology as compared to L2 learners whose L1 orthography is more transparent. This seems to be true for both intermediate and advanced L2 speakers. Although it is difficult to directly compare proficiency levels across studies, it is possible to roughly classify participants into intermediate and high proficiency categories (Table 3).

### Intermediate Proficiency Readers

Studies of intermediate proficiency ESL readers (see Table 3) have focused on adult learners and used the signature manipulations of regularity, frequency, and homophonic/pseudohomophonic status to look at reading procedure. For example, Wang and Koda (2005) compared Chinese and Korean L1 participants on a naming task. Although the Chinese participants were older than the Korean participants, they were well matched in terms of English experience and proficiency. On the naming task, the Chinese L1 participants were less likely to regularize low frequency exception words than were Korean L1 participants. Furthermore, the Korean participants named non-words more accurately. Wang et al. (2003) studied a similar population, but on a different task. They presented participants with a semantic judgment task with four types of foils: similarly spelled homophones, similarly spelled controls, less similarly spelled homophones, and less similarly spelled controls. For example, for the category “type of weather,” the category exemplar was “rain,” the similarly spelled homophone foil was “rein,” and the similarly spelled control was “ruin.” For the category “breakfast food,” the category exemplar

was “cereal,” the less similarly spelled homophone foil was “serial,” and the less similarly spelled control was “several.” Korean participants were more likely to make false alarms to homophone foils whereas Chinese participants were more likely to make false alarms to similarly spelled foils, suggesting that the Korean participants relied more on phonological information whereas Chinese participants relied more on orthographic information during the task (Wang et al., 2003).

Koda (1988) compared native Spanish, Arabic, Japanese<sup>12</sup>, and English speakers on a spelling recognition task (participants see a word and its homophonic foil and are asked to choose the correctly spelled word, e.g., rain, rane) and pseudoword selection task (participants see two non-words and are asked to choose the one that sounds like a real word, e.g., rane, tane). Although all participants were slower on the pseudoword selection task than the spelling recognition task, this difference was most pronounced in the native English speakers and Japanese L1 speakers. The participants also read two passages; in one all of the words were spelled correctly and in the second many words were replaced by their heterographic homophones (e.g., Ted and Bill went hiking in the mountains last weak). The native English speakers and Japanese L1 speakers were slowed to a greater extent on the passage with the heterographic homophones. Furthermore, when the participants were asked to go back and find the homophones, the native English speakers found more than the Japanese L1 speakers, who found more than the Spanish and Arabic L1 speakers.

The other studies examining reading procedure in intermediate level ESL participants have used non-signature manipulations to look at reading procedure. For example, Brown and Haynes (1985) found that, unlike Spanish and Arabic participants, there was no correlation between listening and reading comprehension in Japanese participants. These results suggest that the Spanish and Arabic participants were sounding out the words, and then using their listening comprehension skills to understand the text.

Koda (1990) found that, in contrast to Japanese participants, Arabic and Spanish participants<sup>13</sup> were slowed down when they could not engage in phonological recoding. She gave the participants two passages which described five novel items. In one passage, the names were non-sense pseudowords whereas in the other passage the names were Sanskrit characters that were equally unfamiliar to all participants. The Arabic and Spanish participants were slowed down while reading the passage with the Sanskrit characters, presumably because they were unpronounceable and hence the participants had to rely solely on orthographic information to remember them. In contrast, the Japanese participants read both passages at approximately the same speed. These results can be best explained using the triangle model; if the Japanese participants are accustomed to mainly relying on the orthography to semantics pathway, the Sanskrit characters do not force them to change strategy. In contrast, if

<sup>12</sup>The EFL participants were matched on English ability as measured by cloze and listening comprehension.

<sup>13</sup>All three groups were matched on English ability, as measured by a cloze test.



**TABLE 3 | Criteria used to classify participants as intermediate (italicized) or high (bolded) proficiency.**

Study	TOEFL score	Time living in English-speaking country	No. of years studying English	Self-rated English fluency	Other
<i>Wang and Koda, 2005</i>	~550 (PBT, post-1995)	<1 year	~7	~2/4	Michigan score: ~63
<i>Wang et al., 2003</i>	~560 (PBT, post-1995)	<1 year	~8	~2/4	Michigan score: ~64
<i>Koda, 1990</i>			≥6		~17/31 on a cloze test (Bachman, 1982)
<i>Brown and Haynes, 1985</i>					Enrolled in English L2 classes; presumably low proficiency
<i>Koda, 1988</i>			≥6		~17/31 on a cloze test; ~27/35 on a reading comprehension test
<i>Wade-Woolley, 1999</i>	~50% accuracy on the vocab/reading comp section of TOEFL (PBT, pre-1995)	Japanese participants: ~3 weeks in Canada; Russian participants: 4 years in Israel	~9		Japanese participants in low-intermediate ESL classes; Russian participants identified by the University as having poor English skills; ~79/106 on the Woodcock Word Attack Test
<i>Akamatsu, 2005</i>	Average scaled score of 52.1 on the vocab/reading comp section of TOEFL (PBT, pre-1995)	Most likely never lived in an English-speaking country			Currently undergraduate students in the English department of a Japanese university
<b>Wang and Geva, 2003</b>					~40th percentile on PPVT
<b>Akamatsu, 1999</b>	80–100% accuracy (average scaled score of 63.8) on the vocab/reading comp section of TOEFL, (PBT, pre-1995)				Currently graduate students at a Canadian university
<b>Akamatsu, 2003</b>	>85% accuracy on the vocab/reading comp section of TOEFL (PBT, pre-1995)				Passed reading speed criterion in the Gray Oral Reading Test; currently graduate students or recent PhD graduates at a Canadian university
<b>Perfetti et al., 2007</b>					Described as fluent by the authors
<b>Tan et al., 2003</b>	≥570 (PBT, post-1995)	3–7 years	~12	~4/5	

The most widely used measure of proficiency is the Test of English Language Proficiency (TOEFL). The TOEFL has gone through four major revisions in recent history: (1) before 1995, there was a paper-based test (PBT). It had three sections (vocabulary/reading comprehension, listening comprehension, structure/written expression) (ETS, 1994; Encomium Publications, 2000). The vocabulary/reading comprehension section score could either be expressed as % accuracy or as a scaled score that ranged from 22–67 (ETS, 1994). (2) After 1995, a new PBT was introduced that had three sections (reading comprehension, listening comprehension, structure/written expression). The total scaled score ranged from 310–677 (Encomium Publications, 2000; ETS, 2014). (3) A computer based version was introduced in 1998 and (4) an internet-based version in 2005 (Wall and Horák, 2008).

the Spanish and Arabic participants rely on phonology as well, the Sanskrit characters require a change in strategy.

Wade-Woolley (1999) found that Japanese adults outperform Russian adults on *confrontation spelling tasks*, where they have to decide which of the presented orthographic strings is the correct spelling for the auditorily presented word. This result suggests that the Japanese participants are more likely to read via stored holistic orthographic patterns. One problem with this study is that the participants were not well matched; the Russian participants had been living in Israel for an average of 3.9 years whereas the Japanese participants were in Canada for an average of 3 weeks. However, because the Russian participants had spent more time in an English-speaking country, one would expect them to outperform the Japanese participants. Because the opposite result was found, we can be fairly confident that differing levels of English experience were not confounding the results.

### High Proficiency Readers

It is possible that as people gain proficiency in English, they adjust the weightings given to the two pathways and use a reading procedure more suitable for English's level of transparency.

However, studies using more skilled populations (see **Table 3**) have found no evidence of this and have instead supported the L2 persistency claim, that first language effects can be found even among advanced L2 speakers. Studies of more skilled ESL learners have used the signature manipulation of lexicality as well as unique behavioral paradigms and neuroimaging.

Wang and Geva (2003) compared second grade native Cantonese speakers who grew up in Canada (albeit lived in Cantonese-speaking communities) and began learning English in first grade (when they entered mainstream school) to their native English speaking peers. The groups were not well matched on cognitive measures; the native English speakers had higher vocabulary scores, but the native Cantonese speakers had higher non-verbal reasoning skills. Despite equivalent real word spelling skills, the native English speakers were better at spelling non-words. However, the native Cantonese speakers displayed superior orthographic skills. For example, the native Cantonese children outperformed their native English-speaking peers on a confrontation spelling task, even while controlling for non-verbal reasoning. They also showed higher performance in a task in which they have to reproduce briefly

displayed pronounceable and non-pronounceable non-words from memory. Furthermore, the native Cantonese children were less affected by pronounceability, suggesting that they are less likely to use phonological recoding to help them remember the non-words<sup>14</sup>.

Akamatsu (1999, 2003) looked at the effect of visual distortion (case alternation) on word naming (1999) and passage reading (2003) among Persian, Chinese, and Japanese adults who were well matched in terms of English experience and proficiency. The Chinese and Japanese participants were slowed down to a greater degree by the case alternation than were the Persian participants. Interestingly, the effect of case alternation in the Akamatsu (1999) study was restricted to low frequency words. Because case alternation disrupts word shape cues, this finding suggests that L1 readers of morphosyllabaries strongly rely on word shape cues for all words, whereas L1 readers of alphabets mostly rely on word shape cues for high frequency words.

Up to this point, I have claimed that first language effects on second language reading are robust (i.e., they remain even with increasing second language proficiency) and used cross-study comparisons as evidence for this. However, this evidence is relatively weak, as different tasks were used in every study. One study was able to compare different proficiency levels on the same task and found no effect of proficiency, providing stronger evidence the L2 persistency claim. Akamatsu (2005) ran low proficiency Japanese–English bilinguals on the same task that he used in his 1999 study (which used high proficiency Japanese–English bilinguals) and compared the effect of case alternation on naming in both groups. There was no proficiency by case interaction for either RT or accuracy, suggesting that the effects of case alteration remain constant with increasing L2 proficiency.

Neuroimaging studies have confirmed that Chinese L1 participants tend to read in a “whole word” style, by demonstrating that, even while reading more transparent orthographies, Chinese participants use orthographic brain regions associated with reading morphosyllabaries rather than phonological brain regions associated with reading phonographic systems. For example, Perfetti et al. (2007) found that fluent Chinese–English bilinguals show bilateral activation in posterior visual areas when passively viewing words in both languages, whereas native English speakers show a left-dominant pattern. Tan et al. (2003) found that Chinese–English bilinguals strongly activate the middle frontal cortex when making rhyme judgments for both English and Chinese words. In contrast, the native English speakers used the inferior and frontal superior cortices to a greater degree when making rhyme judgments in English. Together, the results from these studies demonstrate that even when native Chinese speakers who are highly proficient in English are reading in English, they neurally process the visual input in a manner that is more similar to how they process Chinese, rather than how a native English speaker would process the same input.

<sup>14</sup>The main effect and interaction held even while controlling for non-verbal reasoning.

## Instructional Method of L1 Literacy

Not only is the transparency of the L1 orthography important, but so is the instructional method of L1 literacy. In China and Taiwan, children are taught to read using alphabets known as pinyin and Zhu-Yin-Fu-Hao, respectively (Table 2), before being introduced to characters. In contrast, in Hong Kong, children are introduced to characters immediately (Wang and Geva, 2003). Characters are taught using a “look-and-say” method, where children are asked to memorize the meaning and pronunciations of characters, without the mediation of an alphabetic system (Holm and Dodd, 1996).

Although alphabetic orthographies are sometimes used to teach Chinese, skilled adults only read characters. In contrast, skilled readers of Japanese must switch between three different orthographies (kanji, hiragana, and katakana) within the same text (Figure 2). Kanji is morphosyllabic and is typically used for content words. Hiragana and katakana (collectively called kana) are both syllabic. Hiragana is used to represent native Japanese words (such as participles and verb endings) and katakana is used for loan words. When learning Japanese, children learn to read all words (even content words that are typically written in kanji) using kana. Later, kanji characters are slowly introduced.

It is important to note that “differences in instructional method” has a much different meaning for alphabets than for morphosyllabaries. For languages with alphabetic orthographies (French, English), children are only learning one orthography, but the method by which they are taught to read that orthography differs in terms of how much phonics is included. In contrast, for languages with morphosyllabic orthographies (Chinese and Japanese), instructional method primarily refers not to how children are taught to read a given orthography, but to how many orthographies they are taught (although there may also be differences in terms of how much attention is drawn to phonetic radicals). Whether or not children are exposed to an orthography that codes for phonological units early in literacy acquisition may affect their reading procedure when they begin to learn a second language.

Studies examining the effect of instruction on L2 reading procedure have mainly relied on lexicality manipulations, although one non-signature experimental paradigm was used as well. The results from pseudoword reading tasks reviewed below suggest that accuracy differences between L1 readers of morphosyllabaries and alphabets can only be found if the readers

私は東京でコーヒーを飲む

**FIGURE 2 | This Japanese sentence means “I drink coffee in Tokyo” and is pronounced “Watashi wa Tokyo de koohii o nomu.”** The red graphs are written in kanji, the blue graphs are written in hiragana, and the purple graphs are written in katakana. The first kanji graph (red) means “I,” the second and third graphs mean “Tokyo,” and the fourth means “drink.” Note that the kanji characters are used for the content words. The first hiragana graph (blue) is a subject marker, the second is a location marker, the third is an object marker, and the fourth serves to conjugate the verb. The purple graphs mean coffee, pronounced “koohii,” a loan word from English.

of morphosyllabaries had no experience with a phonologically based orthography early in their learning. This may be why conflicting results have been found in the literature. For example, two studies have found no differences between participants with morphosyllabic and alphabetic backgrounds in terms of pseudoword reading accuracy (Koda, 1999; Wade-Woolley, 1999), whereas one study has (Holm and Dodd, 1996). Koda (1999) compared Chinese and Korean participants who were matched in terms of their TOEFL scores and Wade-Woolley (1999) compared Russian and Japanese participants who were matched in terms of their TOEFL scores and word reading. Both studies found that the two studied groups were equally accurate at pseudoword reading. In contrast, when Holm and Dodd (1996) compared students from Vietnam, Mainland China, and Hong Kong who were matched in terms of their real word reading and spelling abilities, they found that students from Hong Kong had significantly lower non-word reading accuracy than the other groups. In this study, a difference was found between students from Hong Kong and those from Mainland China, even though they both read morphosyllabic orthographies, likely because students from Mainland China were taught the alphabetic system of pinyin before they began instruction in Chinese characters, unlike the students from Hong Kong. Perhaps Koda (1999) and Wade-Woolley (1999) were unable to find significant effects because their participants had learned pinyin and kana, respectively, both of which code for phonological units.

Although pseudoword reading accuracy effects have only been demonstrated using participants who have no experience with writing systems that code for phonological units, RT effects have been found using Japanese participants (who have experience with phonologically based kana), perhaps because RT measures are more sensitive than accuracy measures. For example, Brown and Haynes (1985) found that Japanese participants showed a greater lexicality effect on their RT during a word/non-word reading task than either Spanish or Arabic participants did.

In addition to instructional effects on pseudoword reading, instructional effects have also been demonstrated on word reading tasks. Scholfield and Chwo (2005) asked Chinese bilinguals to make meaning similarity judgments in both Chinese (reviewed previously) and in English. They were shown two words, and had to judge whether or not they were semantically related. Some of the word pairs were phonologically similar (e.g., “right,” “write”), whereas others were graphically similar (e.g., “mother,” “bother”). They found that Taiwanese participants were both faster and more accurate on the graphically similar word pairs as compared to the phonologically similar word pairs, whereas the reverse was true for the Hong Kong participants. Although this data is suggestive of a difference in reading procedure, it is important to note that the Hong Kong participants had more weekly English lessons than did the Taiwanese participants. Their greater English fluency was reflected by their faster overall RTs to the English stimuli.

## Extension to Other Languages

In addition to the numerous studies done with English L2 learners, two studies have also looked at Japanese L2 learners.

They have used both behavioral and neurocognitive measures and found that the results were consistent with English L2 studies. Chikamatsu (1996) compared English and Chinese L1 participants who were living in the U. S. and were in the same Japanese as a foreign language class. That class was their first introduction to Japanese, so all of the participants had the same educational experience, both in terms of spoken and written Japanese. Because they were beginners, they had only learned kana and had not been introduced to kanji. Participants performed a lexical decision task with three types of stimuli: familiar words (e.g., native Japanese words written in hiragana), unfamiliar words (e.g., loan words written in hiragana), and non-words. The results were consistent with the EFL research; Chinese L1 participants relied more on orthographic information and less on phonological information than did alphabetic L1 participants. The English and Chinese participants were matched for overall RT. However, the Chinese participants were slowed to a greater degree when switching from the familiar to unfamiliar condition than were the English participants, suggesting that they were using a visual-based strategy. Furthermore, the English participants demonstrated a stronger relationship between word length and RT than did the Chinese participants, suggesting that they were relying more heavily on phonological decoding.

Yokoyama et al. (2013) studied Chinese and Korean L1 readers who had studied Japanese for an average of 2.5 years. They found that when the participants performed a lexical decision task in their L2 orthography, Japanese kana, the Chinese L1 participants activated the left middle frontal gyrus more than the Korean L1 participants did. The left middle frontal gyrus is believed to be a phonological processor for morphosyllabic graphs. These results nicely dovetail with the ESL research, that L1 readers of morphosyllabaries use the same neural mechanisms when reading both their first and second languages.

The L2 research has demonstrated that L1 readers of morphosyllabic orthographies (Chinese, Japanese) tend to use less sublexical phonology than do L1 readers of more transparent orthographies (Korean, Russian, Spanish, Arabic, Persian, Vietnamese) while reading in their second language, especially if they were not introduced to a phonologically based orthography such as pinyin or kana during literacy acquisition. These effects can be explained by the assimilation/accommodation hypothesis (Perfetti et al., 2007), which states that people only change their reading procedure if necessitated by the properties of the L2 orthography. L1 readers of morphosyllabaries are accustomed to heavily weighting the morphological pathway, and because it is possible to read more transparent orthographies using the same reading procedure (even if it is not optimal), they will not change their reading procedure.

These first language effects are persistent; they can be found in beginning, intermediate, and advanced second language learners. The second language orthographies that have been studied are English and Japanese kana. The two orthographies are quite different; English has a moderately opaque alphabet whereas Japanese kana is a transparent syllabary. Because similar effects were found for both orthographies, we can predict that similar effects would be found if more transparent alphabets (e.g., Serbo-Croatian, French) or alphasyllabaries (e.g., Hindi, Thai) were

studied as the second language. However, it remains unclear what effects would be seen if an abjad was chosen as the L2 orthography. Note the interesting difference between L1 and L2 learners: although instructional method has little influence on reading procedure when learning a transparent L1 orthography, language background is able to exert an effect on reading procedure when learning a transparent L2 orthography.

All of the studies reviewed above have compared L1 readers of morphosyllabaries to L1 readers of more transparent orthographies. An interesting area of future research would be to expand this research to other L1 groups. For example, I predict that L1 readers of English would show less reliance on sublexical phonology than would L1 readers of Spanish, German, Portuguese, and French when reading in Japanese kana, Greek, or Russian as their L2 orthography. There is limited evidence to support this hypothesis; Koda (1988) found differences in reading procedure between native English speakers and Spanish–English bilinguals. Furthermore, I hypothesize that English speakers from New Zealand may use less sublexical phonology while learning a second language than English speakers from Scotland because of the instructional differences in those two countries.

Some research has been done comparing L1 abjad (mainly Arabic) to other L1 groups learning English. However, the studies have primarily focused on Arabic speakers' relatively poor word recognition skills, while controlling for other language skills (see Ryan and Meara, 1991; Fender, 2003, 2008b). Although this research is very interesting, it does not answer the main question of this review, specifically what reading procedure people use while reading. Therefore, it would be interesting to expand the work with Arabic L1 readers to look at reading procedure.

## Alternative Views

Although there is significant evidence that L1 readers of morphosyllabaries rely more heavily on the morphological pathway than do L1 readers of other orthographies while reading in their L2, there are some findings that do not neatly fit into that theory. For example, Akamatsu (1999) found that Chinese and Japanese participants showed a greater regularity effect on their RTs and that Chinese participants showed a greater regularity effect on their accuracy than did the Persian participants—the opposite of the expected result. Similarly, Wang and Koda (2005) found that Chinese participants showed a greater regularity effect on their accuracy than did Korean participants. Wang and Koda (2005) were able to account for this unexpected result by doing an error analysis: Korean participants were more likely to regularize irregular words than were Chinese participants. Akamatsu (1999) did not report error types so it is unclear whether the same pattern holds for his study. Therefore, although the research broadly supports a difference in L2 reading procedure based on L1 literacy experiences, there are some anomalous findings.

Yamada (2004) suggested that first language influences on reading procedure may be due, not to the transparency of the orthography, but to the phonological properties of the language itself. If the L2 has a more complex phonological system than the L1, L2 learners may find it difficult to use sublexical phonology and therefore rely on morphological information. English is more

phonologically complex than Chinese, which may be why L1 Chinese speakers use morphological information while reading in English. However, this explanation seems unlikely because differences were found between participants from Hong Kong and Mainland China on L2 English tasks, even though they both spoke a phonologically simple language. Furthermore, Chinese participants differed from both English and Korean participants while learning Japanese, which is also a phonologically simple language. Therefore, the orthographic transparency explanation seems to best account for all of the data.

As reviewed above, Scholfield and Chwo (2005) found differences between Taiwanese and Hong Kong students on an English word decision task. The authors acknowledged that these differences could be due to how Chinese is taught in the two countries (using Zhu-Yin-Fu-Hao in Taiwan and the whole word method in Hong Kong, the hypothesis that was espoused in this review) or to how English is taught in the two countries. During English instruction, Taiwanese schools focus on phonics whereas Hong Kong schools use the whole word method. It is impossible to know whether the differences in reading procedure stem from the manner of L1 or L2 literacy instruction in this case, because the two factors are confounded. The other study which found an effect of L1 instructional method on L2 reading procedure (Holm and Dodd, 1996) did not report on the English instructional methodology in the populations studied, so we do not know whether it was a confounding factor.

Studies examining the effect of L1 literacy *instructional method* on L2 reading procedure have not closely controlled for L2 literacy instructional method, making it impossible to know with certainty whether or not L1 literacy instructional method is sufficient to exert an effect on L2 reading procedure. It is also possible that some of the findings of L1 orthographic *transparency* were also confounded by L2 instructional method. For example, Wade-Woolley (1999) pointed out that, in Japan, English instruction is often very similar to kanji instruction; whole words are presented for memorization. However, some studies have successfully demonstrated the effect of L1 orthographic transparency on L2 reading procedure while controlling for L2 literacy instructional method. For example, Wang and Geva (2003) were able to control for L2 literacy instruction (but not English language experience) by comparing Cantonese–English bilinguals to their native speaking peers. In contrast, Chikamatsu (1996) was able to control for all aspects of L2 experience by comparing students who were in the same introductory Japanese class. Because significant differences in L2 reading procedure were found in both of these studies, it is clear that differences in L1 orthographic transparency are sufficient to affect L2 reading procedure.

## Conclusion

The division of labor between the sublexical phonological and morphological pathways can vary depending on word type, context, and the early literacy experiences of an individual. This review focused on variations across individuals and demonstrated that early literacy experiences, both which

language one first learned to read and how one was taught to read, can have profound and long-lasting impacts on reading procedure. People who learn a more transparent orthography use more sublexical phonology while reading, whereas people who learn a more opaque orthography rely more heavily on morphological/orthographic information. For readers of more opaque orthographies (e.g., English, Chinese), instructional method also impacts reading procedure. These effects are measurable in both beginning and advanced readers.

Not only do early literacy experiences affect how one reads in one's first language, they also affect how one reads in a foreign language. L1 readers of morphosyllabic orthographies use less sublexical phonology than do L1 readers of more transparent orthographies and these effects are measurable in beginning, intermediate, and advanced L2 learners of English and Japanese kana, in children and adults, and in comparison to participants with various L1 backgrounds. However, they may be moderated by whether or not a reader was introduced to an orthography that codes for phonological units (e.g., pinyin, Zhu-Yin-Fu-Hao, kana) during early literacy acquisition.

Although not the focus of this review, it is interesting to consider the clinical implications of differential reading procedures. For example, brain damage can selectively impair either the phonological or semantic pathway. The same patterns of brain damage may differentially affect the severity of reading impairment depending on the dominant reading procedure prior to injury. For example, damage to the semantic pathway would be less damaging in a person who primarily depended on the

phonological pathway than someone who primarily depended on the semantic pathway prior to injury (Plaut et al., 1996). When assessing the impact of selective brain damage, it may be important to consider the person's first language as well as their educational experiences.

It is important to note that that the majority of the conclusions in this review was drawn based on English, which in many regards has an outlier orthography. Where possible, research on other languages was included and hypotheses were made as to how applicable the conclusions drawn from English are to other languages. There is currently much more work being done with other languages, so hopefully some of the hypotheses made in this review can be empirically tested in the near future.

## Acknowledgments

Publication was funded by the National Science Foundation (Award #0354420) through the Pittsburgh Science of Learning Center. This manuscript partially fulfilled the requirements for the comprehensive exam of AB. I would like to thank my committee members, Charles Perfetti, Julie Fiez, and Keiko Koda, for their comments on drafts of this manuscript. I would also like to thank Joseph Stafura and Elizabeth Hirshorn for their comments on portions of this manuscript. Finally, I would like to thank Li-Yun Chang, Xiaoping Fang, Jiexin Gu, Lindsay Harris, Regina Calloway, Alba Tuninetti, Rajeev Bhide, Dnyanada Bhide, Krystina Teoh, Rivka Rosenthal, and Elizabeth Hirshorn for their help illustrating different orthographies.

## References

- Akamatsu, N. (1999). The effects of first language orthographic features on word recognition processing in English as a second language. *Read. Writ.* 11, 381–403. doi: 10.1023/A:1008053520326
- Akamatsu, N. (2003). The effects of first language orthographic features on second language reading in text. *Lang. Learn.* 53, 207–231.
- Akamatsu, N. (2005). "Effects of second language reading proficiency and first language orthography on second language word recognition," in *Second Language Writing Systems*, eds V. J. Cook and B. Bassetti (Tonawanda, NY: Multilingual Matters), 238–259.
- Akita, K., and Hatano, G. (1999). "Learning to read and write in Japanese," in *Learning to Read and Write: A Cross-linguistic Perspective* Vol. 2, eds M. Harris and G. Hatano (Cambridge: Cambridge University Press), 214–234.
- Asfaha, Y. M., Kurvers, J., and Kroon, S. (2009). Grain size in script and teaching: literacy acquisition in Ge'ez and Latin. *Appl. Psycholinguist.* 30, 709–724. doi: 10.1017/S0142716409990087
- Bachman, L. (1982). The trait structure of cloze test scores. *TESOL Q.* 16, 61–70. doi: 10.2307/3586563
- Besner, D., and Smith, M. C. (1992). "Basic processes in reading: Is the orthographic depth hypothesis sinking?" in *Orthography, Phonology, Morphology, and Meaning*, eds R. Frost and L. Katz (Amsterdam: North-Holland), 45–66.
- Bhide, A., Gadgil, S., Zelinsky, C. M., and Perfetti, C. A. (2013). Does reading in an alphasyllabary affect phonemic awareness? *Writ. Syst. Res.* 6, 73–93. doi: 10.1080/17586801.2013.855619
- Brown, G. D., and Deavers, R. P. (1999). Units of analysis in nonword reading: evidence from children and adults. *J. Exp. Child Psychol.* 73, 208–242. doi: 10.1006/jecp.1999.2502
- Brown, T., and Haynes, M. (1985). "Literacy background and reading development in a second language," in *The Development of Reading Skills*, ed. H. Carr (San Francisco: Jossey-Bass), 19–34. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/cd.23219852704/abstract>
- Chikamatsu, N. (1996). The effects of L1 orthography on L2 word recognition: a study of American and Chinese learners of Japanese. *Stud. Second Lang. Acquisit.* 18, 403–432. doi: 10.1017/S0272263100015369
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. C. (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
- Connelly, V., Johnston, R., and Thompson, G. B. (2001). The effect of phonics instruction on the reading comprehension of beginning readers. *Read. Writ.* 14, 423–457. doi: 10.1023/A:1011114724881
- Connelly, V., Thompson, G. B., Fletcher-Flinn, C. M., and McKay, M. F. (2009). "Does the type of reading instruction have an influence on how readers process print," in *Contemporary Perspectives on Reading and Spelling*, eds C. Wood and V. Connelly (New York: Routledge), 239–253.
- Deacon, S. H., Desrochers, A., and Levesque, K. (in press). "The Cross-linguistic study of reading and spelling acquisition: the case of French," in *Reading Acquisition Across Languages and Writing Systems*, eds L. Verhoeven and C. A. Perfetti (Cambridge: Cambridge University Press).
- Defior, S., and Serrano, F. (in press). "Reading acquisition in Spanish," in *Reading Acquisition Across Languages and Writing Systems*, eds L. Verhoeven and C. A. Perfetti (Cambridge: Cambridge University Press).
- Ellis, N. C., and Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: orthographic transparency effects evinced with frequency-matched tests. *Appl. Psycholinguist.* 22, 571–599. doi: 10.1017/S0142716401004052
- Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H. P., Polyzoë, N., et al. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Read. Res. Q.* 39, 438–468. doi: 10.1598/RRQ.39.4.5

- Encomium Publications (2000). *About TOEFL*. Available at: <https://www.encomium.com/webmentor/abouttoefl.html>
- ETS (1994). *TOEFL Test and Score Manual Supplement 1994-95 Edition*. Available at: <https://www.ets.org/Media/Research/pdf/TOEFL-SUM-9495.pdf>
- ETS (2014). *Understanding Your TOEFL® PBT Test Scores*. Available at: <http://www.ets.org/toefl/pbt/scores/understand/>
- Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Appl. Psycholinguist.* 24, 289–315. doi: 10.1017/S014271640300016X
- Fender, M. (2008a). “Arabic literacy development and cross-linguistic effects in subsequent L2 literacy development,” in *Learning to Read Across Languages: Cross-Linguistic Relationships in First-and Second-Language Literacy Development*, eds K. Koda and A. M. Zehler (New York, NY: Routledge), 101–124.
- Fender, M. (2008b). Spelling knowledge and reading development: insights from Arab ESL learners. *Read. Foreign Lang.* 20, 19–42. Available at: <http://nflrc.hawaii.edu/RFL/April2008/fender/fender.html>
- Feng, G., Miller, K., Shu, H., and Zhang, H. (2001). Rowed to recovery: The use of phonological and orthographic information in reading Chinese and English. *J. Exp. Psychol.* 27, 1079–1100. doi: 10.1037//0278-7393.27.4.1079
- Foorman, B. R., Francis, D. J., Novy, D. M., and Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *J. Educ. Psychol.* 83, 456–469. doi: 10.1037//0022-0663.83.4.456
- Frost, R. (2012). Towards a universal model of reading. *Behav. Brain Sci.* 35, 1–17. doi: 10.1017/S0140525X11001841
- Frost, R., and Bentin, S. (1992). *Reading Consonants and Guessing Vowels: Visual Word Recognition in Hebrew Orthography*. New Haven, CT: Haskins Laboratory.
- Frost, R., Katz, L., and Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 104–115. doi: 10.1037/0096-1523.13.1.104
- Goswami, U., Ziegler, J. C., Dalton, L., and Schneider, W. (2001). Pseudohomophone effects and phonological recoding procedures in reading development in English and German. *J. Mem. Lang.* 45, 648–664. doi: 10.1006/jmla.2001.2790
- Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev.* 111, 662–720. doi: 10.1037/0033-295X.111.3.662
- Hirshorn, E., Durisko, C., Moore, M., Perfetti, C. A., and Fiez, J. (2015). Effects of mapping principle on the learning trajectory of two artificial orthographies. *Paper Presented at the Annual Meeting of the Society for the Scientific Study of Reading, Kohala Coast, Hawaii*.
- Holm, A., and Dodd, B. (1996). The effect of first written language on the acquisition of English literacy. *Cognition* 59, 119–147. doi: 10.1016/0010-0277(95)00691-5
- Johnston, R. S., and Thompson, G. B. (1989). Is dependence on phonological information in children's reading a product of instructional approach? *J. Exp. Child Psychol.* 48, 131–145. doi: 10.1016/0022-0965(89)90044-1
- Johnston, R. S., Thompson, G. B., Fletcher-Flinn, C. M., and Holligan, C. (1995). The functions of phonology in the acquisition of reading: lexical and sentence processing. *Mem. Cogn.* 23, 749–766. doi: 10.3758/BF03200927
- Kang, J. S. (1993). “Analysis of semantics of semantic–phonetics compound characters in modern Chinese,” in *Information Analysis of Usage of Characters in Modern Chinese*, ed. Y. Chen (Shanghai: Shanghai Education Publisher).
- Katz, L., and Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *J. Exp. Psychol.* 9, 157–166.
- Katz, L., and Frost, R. (1992). “The reading process is different for different orthographies: the orthographic depth hypothesis,” in *Speech Research Status Report. Haskins Laboratory*, ed. C. A. Fowler (Brazil: Elsevier). Available at: [http://www.sciencedirect.com/science/article/pii/S0166411508627892/pdf?md5=b61514fbbf905b698fe18008dc211d4b&pid=1-s2.0-S0166411508627892-main.pdf&\\_valck=1](http://www.sciencedirect.com/science/article/pii/S0166411508627892/pdf?md5=b61514fbbf905b698fe18008dc211d4b&pid=1-s2.0-S0166411508627892-main.pdf&_valck=1)
- Koda, K. (1988). Cognitive process in second language reading: transfer of L1 reading skills and strategies. *Second Lang. Res.* 4, 133–155. doi: 10.1177/026765838800400203
- Koda, K. (1990). The Use of L1 reading strategies in L2 reading: effects of L1 orthographic structures in L2 phonological recoding strategies. *Stud. Second Lang. Acquisit.* 12, 393–410. doi: 10.1017/S0272263100009499
- Koda, K. (1999). Development of L2 intraword orthographic sensitivity and decoding skills. *Mod. Lang. J.* 83, 51–64. doi: 10.1111/0026-7902.00005
- Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., and Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Read. Res. Q.* 48, 61–76. doi: 10.1002/rrq.038
- Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *Eur. J. Psychol. Educ.* 15, 239–257. doi: 10.1007/BF03173177
- Landerl, K. (in press). “Written language German,” in *Reading Acquisition Across Languages and Writing Systems*, eds L. Verhoeven and C. A. Perfetti (Cambridge: Cambridge University Press).
- Leybaert, J., and Content, A. (1995). Reading and spelling acquisition in two different teaching methods: a test of the independence hypothesis. *Read. Writ.* 7, 65–88. doi: 10.1007/BF01026948
- Li, T., and McBride-Chang, C. (2014). “How character reading can be different from word reading in Chinese and why it matters for Chinese reading development,” in *Reading Development and Difficulties in Monolingual and Bilingual Chinese Children*, eds X. Chen, Q. Wang, and Y. C. Luo (Berlin: Springer), 49–65. doi: 10.1007/978-94-007-7380-6\_3
- MacWhinney, B. (2001). “The competition model: the input, the context, and the brain,” in *Cognition and Second Language Instruction*, ed. P. Robinson (New York: Cambridge University Press), 69–90.
- Nag, S. (2011). “The akshara languages: what do they tell us about children's literacy learning?,” in *Language-Cognition Interface: State of the Art*, eds R. K. Mishra and N. Srinivasan (Munich: Lincom Publishers), 291–310.
- Norris, D., and Brown, G. (1985). Race models and analogy theories: a dead heart? Reply to Seidenberg. *Cognition* 20, 155–168. doi: 10.1016/0010-0277(85)90051-4
- 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
- Perfetti, C. A., and Harris, L. N. (2013). Reading universals are modulated by language and writing system. *Lang. Learn. Dev.* 9, 296–316. doi: 10.1080/15475441.2013.813828
- Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., and Tan, L.-H. (2007). Reading in two writing systems: Accommodation and assimilation of the brain's reading network. *Bilingualism* 10, 131–146. doi: 10.1017/S136672890702891
- Perfetti, C. A., and Zhang, S. (1991). Phonological processes in reading Chinese characters. *J. Exp. Psychol.* 17, 633–643. doi: 10.1037/0278-7393.17.4.633
- Perfetti, C. A., Zhang, S., and Berent, I. (1992). “Reading in English and Chinese: evidence for a “universal” phonological principle,” in *Orthography, Phonology, Morphology, and Meaning*, eds R. Frost and L. Katz (Amsterdam: North-Holland), 227–248. Available at: [http://www.sciencedirect.com/science/article/pii/S0166411508627983/pdf?md5=943bbe68046813255afa4a236f158abf&pid=1-s2.0-S0166411508627983-main.pdf&\\_valck=1](http://www.sciencedirect.com/science/article/pii/S0166411508627983/pdf?md5=943bbe68046813255afa4a236f158abf&pid=1-s2.0-S0166411508627983-main.pdf&_valck=1)
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115. doi: 10.1037/0033-295X.103.1.56
- Ryan, A., and Meara, P. (1991). The case of invisible vowels: Arabic speakers reading English words. *Read. Foreign Lang.* 7, 531–540.
- Saiegh-haddad, E. (in press). “Written Arabic: language, orthography, and reading acquisition,” in *Reading Acquisition Across Languages and Writing Systems*, eds L. Verhoeven and C. A. Perfetti (Cambridge: Cambridge University Press).
- Scholfield, P., and Chwo, G. S.-M. (2005). “Are the L1 and L2 word reading processes affected more by writing system or instruction?,” in *Second Language Writing Systems*, eds V. J. Cook and B. Bassetti (Tonawanda, NY: Multilingual Matters), 215–237.
- Seidenberg, M. S., and McClelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568. doi: 10.1037/0033-295X.96.4.523

- Spencer, L. H., and Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *Br. J. Psychol.* 94, 1–28. doi: 10.1348/000712603762842075
- Tan, L. H., Spinks, J. A., Feng, C.-M., Siok, W. T., Perfetti, C. A., Xiong, J., et al. (2003). Neural systems of second language reading are shaped by native language. *Hum. Brain Mapp.* 18, 158–166. doi: 10.1002/hbm.10089
- Taylor, I., and Taylor, M. M. (1995). *Writing and Literacy in Chinese, Korean and Japanese*, Vol. 3. Philadelphia, PA: John Benjamins Publishing.
- Taylor, J. S. H., Davis, M. H., and Rastle, K. (2015). Learning to read aloud versus learning to comprehend: the impact of training focus on behavioural and neural responses to artificial scripts. *Paper Presented at the Annual Meeting of the Society for the Scientific Study of Reading, Kohala Coast, Hawaii.*
- Thompson, G. B., Connelly, V., Fletcher-Flinn, C. M., and Hodson, S. J. (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Mem. Cogn.* 37, 223–234. doi: 10.3758/MC.37.2.223
- Thompson, G. B., and Johnston, R. S. (2000). Are nonword and other phonological deficits indicative of a failed reading process? *Read. Writ.* 12, 63–97. doi: 10.1023/A:1008048712877
- Thompson, G. B., McKay, M. F., Fletcher-Flinn, C. M., Connelly, V., Kaa, R. T., and Ewing, J. (2008). Do children who acquire word reading without explicit phonics employ compensatory learning? Issues of phonological recoding, lexical orthography, and fluency. *Read. Writ.* 21, 505–537. doi: 10.1007/s11145-007-9075-9
- Wade-Woolley, L. (1999). First language influences on second language word reading: all roads lead to Rome. *Lang. Learn.* 49, 447–471. doi: 10.1111/0023-8333.00096
- Wall, D., and Horák, T. (2008). *The Impact of Changes in the TOEFL Examination on Teaching and Learning in Central and Eastern Europe: Phase 2, Coping With Change*. Available at: <http://www.ets.org/Media/Research/pdf/RR-08-37.pdf>
- Wang, M., and Geva, E. (2003). Spelling performance of Chinese children using English as a second language: lexical and visual-orthographic processes. *Appl. Psycholinguist.* 24, 1–25. doi: 10.1017/S0142716403000018
- Wang, M., and Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Lang. Learn.* 55, 71–98. doi: 10.1111/j.0023-8333.2005.00290.x
- Wang, M., Koda, K., and Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: a comparison of Korean and Chinese English L2 learners. *Cognition* 87, 129–149. doi: 10.1016/S0
- Wimmer, H., and Goswami, U. (1994). The influence of orthographic consistency on reading development: word recognition in English and German children. *Cognition* 51, 91–103. doi: 10.1016/0010-0277(94)90010-8
- Yamada, J. (2004). An L1-script-transfer-effect fallacy: a rejoinder to Wang et al. (2003). *Cognition* 93, 127–132. doi: 10.1016/j.cognition.2003.09.014
- Yang, J., McCandliss, B. D., Shu, H., and Zevin, J. D. (2008). “Division of labor between semantics and phonology in normal and disordered reading development across languages,” in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Washington, DC. Available at: <http://www.sacklerinstitute.org/cornell/people/jason.zevin/publications/publications/fp501-yang.pdf>
- Yang, J., Zevin, J. D., Shu, H., McCandliss, B. D., and Li, P. (2006). “A “triangle model” of Chinese reading,” in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Mahwah, NJ, 912–917. Available at: <http://psychling.psych.ac.cn/publications/yangjf/2006YangCogSci.pdf>
- Yokoyama, S., Kim, J., Uchida, S., Miyamoto, T., Yoshimoto, K., and Kawashima, R. (2013). Cross-linguistic influence of first language writing systems on brain responses to second language word reading in late bilinguals. *Brain Behav.* 3, 525–531. doi: 10.1002/brb3.153
- Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3
- Ziegler, J. C., and Goswami, U. (2006). Becoming literate in different languages: similar problems, different solutions. *Dev. Sci.* 9, 429–436. doi: 10.1111/j.1467-7687.2006.00509.x

**Conflict of Interest Statement:** 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 © 2015 Bhide. 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) or licensor 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.



# Polish pseudo-words list: dataset of 3023 stimuli with competent judges' ratings

Kamil K. Imbir<sup>1,2\*</sup>, Tomasz Spustek<sup>3</sup> and Jarosław Żygierewicz<sup>3</sup>

<sup>1</sup> Faculty of Psychology, University of Warsaw, Warsaw, Poland, <sup>2</sup> The Maria Grzegorzewska University, Warsaw, Poland,

<sup>3</sup> Faculty of Physics, University of Warsaw, Warsaw, Poland

Pseudo-words are stimuli, which are useful in research concerning lexical processing. As in the case of existing words, they are language dependent; thus, they should be generated for each language separately. The Polish Pseudo-words List (PPwL) is a dataset presenting a set of 3023 stimuli (words of 4–13 letters long). They were generated using an algorithm substituting random letters in existing words with respect to the frequency of letters in certain positions. We put out the raw set for a competent judges' assessment and included the responses in the dataset. PPwL allows the choice of suitable control stimuli for experiments concerning lexical processing.

## OPEN ACCESS

### Edited by:

Shelia Kennison,  
Oklahoma State University, USA

### Reviewed by:

Mike Bowers,  
University of Maryland School  
of Medicine, USA  
Pawel Mandera,  
Ghent University, Belgium

### \*Correspondence:

Kamil K. Imbir,  
Faculty of Psychology, University  
of Warsaw, 5/7 Stawki Street,  
00-183 Warsaw, Poland  
kamil.imbir@gmail.com

### Specialty section:

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

**Received:** 08 June 2015

**Accepted:** 01 September 2015

**Published:** 15 September 2015

### Citation:

Imbir KK, Spustek T  
and Żygierewicz J (2015) Polish  
pseudo-words list: dataset of 3023  
stimuli with competent judges'  
ratings. *Front. Psychol.* 6:1395.  
doi: 10.3389/fpsyg.2015.01395

**Keywords:** pseudo-words, lexical stimuli dataset, polish language, lexical decision task, lexical processing

## Introduction

Studies in psychology concerning language processing in so-called lexical decision tasks (c.f. Meyer and Schvaneveldt, 1971) require both words of well-known properties (c.f. Imbir, 2014) and pseudo-word stimuli (e.g., Simos et al., 2002; Keuleers and Brysbaert, 2010) following some orthographical and structural rules (c.f. judging procedure). They should especially respect the phonotactic restrictions of a certain language; thus, each needs their own pseudo-word stimuli, respecting language's specificity. Pseudo-word stimuli have no meaning in the lexicon, but it is possible that such stimuli could potentially be a part of the language. Using proper stimuli is especially important when processing differences are measured in EEG paradigms (c.f. Kanske and Kotz, 2007; Barber et al., 2013; Palazova et al., 2013; Imbir et al., submitted) that are sensitive to subtle differences in stimuli classes. Pseudo-words are more complex forms of stimuli than logatomes or non-sense syllables both of which are composed of single syllables. For that reason, to create them, we may use existing syllables as well as artificial (yet pronounceable) ones. Although in the literature the machine pseudo-words generation method exists (c.f. Keuleers and Brysbaert, 2010), at the moment of beginning of our project "Wuggy" generator was not customized to Polish language. For that reason we decided to generate pseudo-words in a different random fashion (but respecting letters probability of occurrence on certain position in certain neighborhood) and then put all of them into judging procedure.

The aim to create the presented dataset was to provide a set of stimuli (varying with degree of fulfillment of the criteria of ideal pseudo-word) for experimental samples in Polish language. To make use of pseudo-word stimuli easier for other researchers, we decided to share our dataset of 3023 pseudo-words. We hope that this will stimulate research on lexical processing in studies using the Polish language. This could lead to a better understanding of word processing in diverse languages.



## Materials and Methods

### Pseudo-Word Generation

The generation procedure engaged two steps. At first, we chose 540 random nouns from a normative database of 4905 Polish words (Imbir, submitted). We wanted them to cover words of different lengths (number of letters ranged from 4 to 13). Then, for each noun chosen, six machine-generated pseudo-words were constructed by substituting randomly selected letters for other letters. These other letters retained their type – vowel or consonant – and had to be one of the three most probable to occur after the preceding and before the successive letter. In fact most probable letter (or letters – the algorithm was random so the same letter position could have been chosen twice or more and thus generate two different pseudo-words) for certain, randomly chosen position was placed instead of original letter. The original letter at chosen position was excluded, so if that letter was most probable to occur at certain position, algorithm replaced it by second most probable. Also if generated stimulus was the same as previously generated (or other existing word included in 4905 words list) algorithm searched for another pseudo-word in order to replace this one. As reference point for probability of occurrence we used whole 4905 word list (Imbir, submitted) representing large number of words from Polish language. The rationale for this choice was expectation that generated pseudo-words should match as much as possible to available lexical stimuli (whole list). Unfortunately, this procedure does not guarantee that pseudo-words respect the phonotactic restrictions of the language, thus further judge competent engagement was crucial. For words of 4–6 letters long, one letter was substituted; for words of 7–9 letters, two letters were substituted, and for words 10–13 letters long, three letters were substituted. In this way, we obtained a list of 3240 pseudo-words.

### Judging Procedure

The third step was to evaluate our list in terms of subjective fulfillment of criteria for pseudo-word stimuli by using competent judges. Pseudo-words were defined as verbal stimuli that (1) are constructed from existing or potential syllables, (2) are possible to read fluently, (3) comply with Polish spelling rules, (4) do not occur in the real language, and (5) do not associate easily with other existing words in the language. We asked five native Polish language speakers

(women), who were students of social science and humanities (including departments for language and literature) to evaluate the whole set of 3240 stimuli and remove those items from the list that did not conform to all of the criteria. After this validation, we inspected the judges' congruency concerning individual pseudo-word stimuli. The advantage of presented methodology is that we asked judges to exclude pseudo-words that can be easily associated with existing words or hard to read and present instant list. Judging is still often needed in case of other pseudo-words stimuli generation procedures.

Eight hundred seventy pseudo-words were positively verified by all five judges and received a congruency index of 1. Next, 988 stimuli were chosen by four of the five judges at the same time (congruency index = 0.8). A total of 537 pseudo-words were indicated by three of the five judges as good stimuli with a congruency index of 0.6. Two judges agreed in the case of 341 stimuli (congruency index = 0.4); for 287 stimuli, only one judge indicated that they were good pseudo-words, while the other four crossed these stimuli out (congruency index = 0.2). Two hundred seventeen (6.7% of initial number) stimuli were excluded by all of competent judges.

### Dataset Description

The Polish Pseudo-words List (PPwL) dataset is deposited at <http://figshare.com/s/1089daa40de311e589a806ec4b8d1f61> and consists of a single xlsx spreadsheet. Pseudo-words are listed in the first column. In the next two columns, one can find the agreement ratio for every single pseudo-word as well as the number of judges indicating that the certain stimuli is a good example of a pseudo-word (max = 5). In the last column, stimulus length is presented as the number of letters in the string. We may assume that the number of 870 pseudo-words with maximum judges congruency represent stimuli of very good quality adhering to five criteria listed above.

### Funding

The project was funded by the National Science Center on the basis of decision DEC: DEC-2013/09/B/HS6/00303.

## References

- Barber, H. A., Otten, L. J., Kousta, S. T., and Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain Lang.* 125, 47–53. doi: 10.1016/j.bandl.2013.01.005
- Imbir, K. (2014). Affective norms for 1,586 Polish words (ANPW): duality of mind approach. *Behav. Res. Methods* 47, 860–870. doi: 10.3758/s13428-014-0509-4
- Kanske, P., and Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res.* 1148, 138–148. doi: 10.1016/j.brainres.2007.02.044
- Keuleers, E., and Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behav. Res. Methods* 42, 627–633. doi: 10.3758/BRM.42.3.627
- Meyer, D. E., and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* 90, 227–234. doi: 10.1037/h0031564
- Palazova, M., Sommer, W., and Schacht, A. (2013). Interplay of emotional valence and concreteness in word processing: an event-related potential study with verbs. *Brain Lang.* 125, 264–271. doi: 10.1016/j.bandl.2013.02.008
- Simos, P. G., Breier, J. I., Fletcher, J. M., Foorman, B. R., Castillo, E. M., and Papanicolaou, A. C. (2002). Brain mechanisms for reading words

and pseudowords: an integrated approach. *Cereb. Cortex* 12, 297–305. doi: 10.1093/cercor/12.3.297

**Conflict of Interest Statement:** 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 © 2015 Imbir, Spustek and Żygierewicz. 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) or licensor 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.*

# Co-lateralized bilingual mechanisms for reading in single and dual language contexts: evidence from visual half-field processing of action words in proficient bilinguals

OPEN ACCESS

**Edited by:**

Shelia Kennison,  
Oklahoma State University, USA

**Reviewed by:**

Gang Peng,  
The Chinese University of Hong Kong,  
Hong Kong  
Hsu-Wen Huang,  
National Taiwan Normal University,  
Taiwan

**\*Correspondence:**

Gregory Króliczak,  
Laboratorium Badania Działań i  
Poznania, Instytut Psychologii, Adam  
Mickiewicz University in Poznań, ul.  
Szamarzewskiego 89B, 60-568  
Poznań, Poland  
krolgreg@amu.edu.pl;  
krol.greg@gmail.com

**Specialty section:**

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

**Received:** 19 February 2015

**Accepted:** 24 July 2015

**Published:** 07 August 2015

**Citation:**

Krefta M, Michałowski B, Kowalczyk J  
and Króliczak G (2015) Co-lateralized  
bilingual mechanisms for reading in  
single and dual language contexts:  
evidence from visual half-field  
processing of action words in  
proficient bilinguals.  
Front. Psychol. 6:1159.  
doi: 10.3389/fpsyg.2015.01159

**Marlena Krefta<sup>1</sup>, Bartosz Michałowski<sup>1,2</sup>, Jacek Kowalczyk<sup>2</sup> and Gregory Króliczak<sup>1\*</sup>**

<sup>1</sup> Action and Cognition Laboratory, Department of Social Sciences, Institute of Psychology, Adam Mickiewicz University in Poznań, Poznań, Poland, <sup>2</sup> Faculty of English, Adam Mickiewicz University in Poznań, Poznań, Poland

When reading, proficient bilinguals seem to engage the same cognitive circuits regardless of the language in use. Yet, whether or not such “bilingual” mechanisms would be lateralized in the same way in distinct—single or dual—language contexts is a question for debate. To fill this gap, we tested 18 highly proficient Polish (L1)—English (L2) childhood bilinguals whose task was to read aloud one of the two laterally presented action verbs, one stimulus per visual half field. While in the single-language blocks only L1 or L2 words were shown, in the subsequent mixed-language blocks words from both languages were concurrently displayed. All stimuli were presented for 217 ms followed by masks in which letters were replaced with hash marks. Since in non-simultaneous bilinguals the control of language, skilled actions (including reading), and representations of action concepts are typically left lateralized, the vast majority of our participants showed the expected, significant right visual field advantage for L1 and L2, both for accuracy and response times. The observed effects were nevertheless associated with substantial variability in the strength of the lateralization of the mechanisms involved. Moreover, although it could be predicted that participants’ performance should be better in a single-language context, accuracy was significantly higher and response times were significantly shorter in a dual-language context, irrespective of the language tested. Finally, for both accuracy and response times, there were significant positive correlations between the laterality indices (LIs) of both languages independent of the context, with a significantly greater left-sided advantage for L1 vs. L2 in the mixed-language blocks, based on LIs calculated for response times. Thus, despite similar representations of the two languages in the bilingual brain, these results also point to the functional separation of L1 and L2 in the dual-language context.

**Keywords:** bilingualism, language context, overt reading, visual half fields, lateralization

## Introduction

In the majority of people, the left hemisphere is typically involved in the control of language and its many related skills. Yet a strength, and in some cases even a direction, of their laterality is often modulated by the actual linguistic experience, including the onset of exposure to different languages and the achieved fluency (e.g., Perani et al., 1998; Klein et al., 2006; Grossi et al., 2010). Indeed, the overall organization of languages in the human brain seems to depend on whether they are acquired simultaneously, or rather the non-native language(s) is (are) acquired later in life, with a degree to which the level of proficiency affects language laterality being a more debatable factor (for a meta-analysis of behavioral studies on bilingual language lateralization, see Hull and Vaid, 2007; for a targeted review of neuroimaging work on this topic, see Abutalebi, 2008).

While most of the studies on language laterality in the bilingual, or multilingual, brain have capitalized on selected aspects of language production (e.g., picture naming or other stimulus-driven word generation) or language comprehension (e.g., semantic categorization of the visually or aurally presented words), relatively little is known about the lateralization of *bilingual mechanisms* involved in such a highly automated linguistic skill as overt reading. Although there is evidence that when a person becomes equally proficient in two or more languages, skilled reading in each of them could engage largely the same neural areas or circuits involved in related mechanisms (cf. Meschyan and Hernandez, 2006; e.g., Parker Jones et al., 2011), this principle might be particularly relevant to situations where two languages either are, even if unintentionally, or must be available for task performance at the very same time (cf. Grosjean, 2001). Consequently, a question remains whether or not the same rule applies when one uses a single language at a given time, and there is neither need nor point to have the other language in readiness (for a brief review, see Wu and Thierry, 2010; see also Van Heuven and Dijkstra, 2010; Spalek et al., 2014).

To shed some light on this issue, we asked proficient Polish-English bilinguals to read aloud action words in one of the two languages alone or—in the later test—to read these same words in the dual-language context. Although such tasks seem quite basic for these two alphabetic scripts, they may still involve many of the left-lateralized mechanisms. This is definitely the case for simple graphic processing of visual word forms (which is typically carried out by the left cortical and subcortical structures, e.g., McCandliss et al., 2003; Cohen and Dehaene, 2004), but the engagement of the dominant hemisphere can be weakened at the level of phonological/semantic processing, depending on the language involved and the age of its acquisition (Leonard et al., 2010; Peng and Wang, 2011; see also Hull and Vaid, 2007). Notably, the relative contribution of the two hemispheres to overt reading should be easily revealed by the pattern of accuracy and/or response times to target words presented in one of the two visual half fields (VHFs). Indeed, when used properly, the method we adopted here is a very reliable measure of cerebral language dominance. Since the outcomes obtained this way have been shown to strongly correlate with neuroimaging results concerning language laterality (Hunter and Brysbaert, 2008), this

method can be successfully used, as a much more economical alternative to the traditional methods, to assess the laterality of the two languages in question.

In sum, this study utilized a very simple but reliable test of language lateralization and applied it to a population of proficient bilinguals. We focused on one particular category of stimuli, i.e., action words, which typically engage concepts that are strongly left lateralized (for review, see Binkofski and Buxbaum, 2013). Therefore, any alleviation of the strength of their lateralized processing could point to a reorganization of the language circuits due to early acquisition of the second language. Moreover, the study involved two separate phases. In the first one, the testing procedures unambiguously pointed to one language only, whereas the second phase invoked the two languages simultaneously. As a result, reading in the single-language context in the VHF paradigm should unequivocally inform us about the laterality of each of the languages. The dual-language context, on the other hand, allowed us to resolve the issue of whether or not the earlier results concerning the laterality of a given language could be affected by the participants' adoption of an intermediate strategy to be equally efficient in both languages, or rather by the between-language interference (or lack of thereof) from the non-target visual field.

Because very proficient bilinguals were tested, we did not expect any differences in response accuracy between the two languages. Yet, if any between-language interference was present, it was more likely to occur in the non-dominant VHF, and possibly for the non-native language. Such effects were predicted unless participants adopted a truly intermediate strategy, which was likely in our highly proficient sample. Finally, given that two of our participants could potentially be classified as infant bilinguals, three others were really close to the adult bilingual category, and the remaining 13 started acquiring the second language between the ages of 7 and 10, we expected a large variability in the strength of the lateralization of their two languages (e.g., Hull and Vaid, 2007). Such variability is an asset (see Biduła and Króliczak, 2015), because it is paramount in testing for correlations between the laterality indices obtained for the two languages. They were of course expected to correlate quite strongly.

## Methods

The first author obtained a positive opinion about the to-be-used procedures and protocols from the local Ethics Committee for Research Involving Human Subjects. Carried out in *Action and Cognition Laboratory* in the Institute of Psychology at Adam Mickiewicz University in Poznań, Poland, the study conformed to the 2013 WMA Declaration of Helsinki.

## Participants

Eighteen healthy volunteers (16 women, age: 18–32, mean = 23.3,  $SD = 2.9$ ) took part in the experiment after giving their written informed consent. All of them had normal or corrected-to-normal visual acuity. Fifteen individuals declared themselves as right-handers, and three as left-handers. All participants were native speakers of Polish (L1) who began to learn English (L2)

as a foreign language between the ages of 5 and 11 (mean = 8.2,  $SD = 2.1$ ). At the time of the experiment, all subjects were highly proficient users of both languages. Their fluency in L2 was established in two ways: on the basis of their field of study—English Philology at Adam Mickiewicz University in Poznań, Poland—and/or the language certificates obtained by passing at some point of their studies standardized tests of English language proficiency, i.e., possessing at least the *Certificate in Advanced English* (CAE), or *International English Language Testing System* (IELTS) with the result of seven points or above.

## Stimuli

Forty Polish and 40 English verbs denoting manual activities that require the use of simple or complex tools were used as stimuli. All the activities were commonly known and frequently performed. This was established in an earlier pilot study, wherein eight individuals rated the familiarity of Polish and English words from a greater set on a scale of 1 (unfamiliar word) to 5 (very familiar word). Only words that received an average of 3 points or above were included in the experimental set. Care was taken to ensure that the verbs in both languages corresponded to each other in their meaning. The stimuli were in their infinitive form (Polish, English), or non-finite, gerund form (English). The rationale for the latter manipulation was to minimize the difference in length between Polish and English verbs, as Polish verbs are typically longer than the English ones. Ten English verbs were kept in their infinitive form to match the shortest Polish verbs. The two sets of words did not differ significantly in terms of the average word length [ $t_{(78)}=0.88$ ,  $p = 0.38$ ]. The number of words starting with voiced or voiceless initial phoneme was the same for both languages, with 18 words beginning with a voiced phoneme and 22 with a voiceless one. For the list of stimuli used in the experiment, see Appendix 1 in Supplementary Materials.

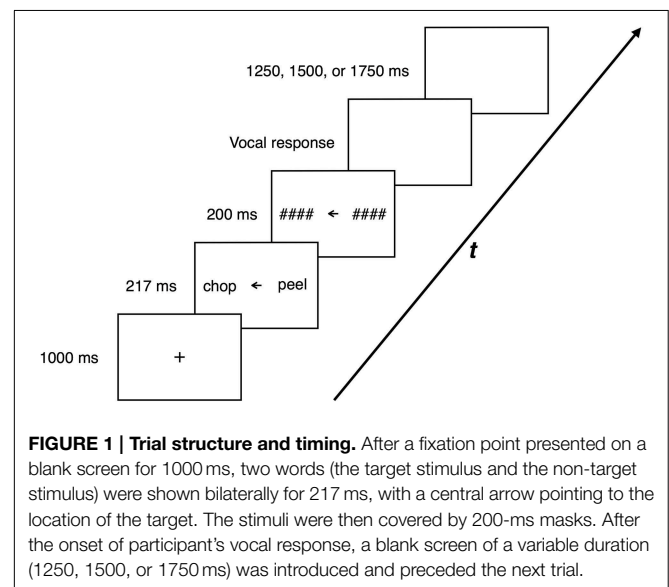
## Procedure

Participants were seated in front of the screen at a viewing distance of ~57 cm. Each trial began with a central fixation cross of 1000-ms duration. Next, two words were presented in the left and right visual field with a central arrow pointing to the left or right. The role of the arrow was to indicate the target word. Participants were instructed to read the target word aloud, and to ignore the other, non-target word. All stimuli were presented on a white background in Arial font, color black, size 50 points, 2° of the visual angle from the central arrow. Although Hunter and Brysbaert (2008) suggested that in a VHF paradigm the stimuli should not be visible for more than 200 ms, our pilot study revealed that with the adopted parameters of the procedure and stimuli, average response accuracy in the dominant field was only about 70%. By using results from a 3-down-1-up staircase procedure, we adjusted the duration of the target stimulus to 217 ms in order to achieve accuracy of approximately 75% (cf. McNair and Harris, 2012). Thus, after 217 ms, both words were masked with strings of hash marks. The length of the presented string was always equal to the length of the masked word. Then, a blank screen appeared and stayed until a vocal response was registered. The response time, as measured by the onset of the

vocal reaction (detected by the SV-1 Smart Voice Key: <http://www.cedrus.com/sv1/>), was recorded by the software used for stimulus presentation (SuperLab 4.5 by Cedrus: <http://www.superlab.com/>). The accuracy of the response was constantly monitored by the experimenter. A blank screen of variable (1250, 1500, or 1750 ms) duration was introduced between the successive trials. The trial structure is depicted in **Figure 1**.

Before the experiment proper, a training session consisting of two single-language blocks, each containing five trials, was administered. Words used during the training session did not appear in the subsequent experimental session. For each participant, the language of the first training block was the same as the language of the first single-language experimental block. The language of instructions always corresponded to the language used in a given block. In the dual-language blocks, the language of instructions was changed every consecutive sentence.

The experiment consisted of six blocks of pseudo-randomly presented trials. At the beginning of each block, participants were informed of its language and/or type (Polish single-language, English single-language, or mixed-language). In the four single-language blocks (two Polish blocks, and two English blocks, 40 trials in each), two words presented in every trial came from the same language (Polish, or English, respectively). In the two mixed-language blocks (80 trials in each), the target word came from one language, and the non-target word came from the other one. In both types of blocks, the primary criterion of assigning words into target–non-target pairs was their length. Each of the eighty stimulus words was presented as a target only four times: two times in single-language blocks (once in the LVE, and once in the RVF), and two times in mixed-language blocks (again, once in the LVE, and once in the RVF). Moreover, in the whole experiment, every word was presented four times as a non-target stimulus. As a result, there were a few trials in which the presented words differed in length by no more than two characters. Mixed-language blocks were always presented



as the last, whereas the order of single-language blocks (two consecutive Polish blocks, and two consecutive English blocks) was counterbalanced across participants.

## Statistical Analyses

The pattern of performance (i.e., accuracy and response times) demonstrated by the three left-handed individuals closely resembled the outcomes of right-handed participants, which is in line with the observation that in the majority of left-handers, language skills are represented in a way similar to their representations in typical, right-handed subjects, at least in the case of simple verbal fluency tests (e.g., Knecht et al., 2000; Króliczak et al., 2011). Therefore, in order to increase statistical power, the results of all 18 participants were analyzed together. To this end, we used two separate repeated-measures Analyses of Variance (ANOVAs), one for accuracy and one for response times to correctly read words. The within-subjects factors were *block type* (single-language, mixed-language), *target language* (Polish, English), and *target location* (left, right). The adopted level of significance was  $\alpha = 0.05$ . If necessary, the required *post-hoc* tests were Bonferroni corrected. Response times exceeding 2.5 s were removed due to the possibility of (1) participants guessing the answer, and/or (2) an equipment malfunction. Also, for reaction times accompanying correctly read words, outliers greater than two standard deviations above or below the mean (calculated for each condition) were removed. Consistent with Hunter and Brysbaert (2008), in such a difficult task and for different reasons (primarily incorrect or too long responses), an average of 34.8% trials for each participant were removed, with only 24.4% trials for target word presented on the right, and 45.2% of trials for target word presented on the left.

In order to determine the hemispheric dominance for the first (Polish) and second (English) language, lateralization indices (LIs) for both languages, within each context (single-language, dual-language), as well as across both tested contexts, were derived through the following formulas, separately for reading accuracy ( $LI_{ACC}$ ) and response times ( $LI_{RT}$ ):

$$LI_{ACC} = [(R - L)/(R + L)] * 100$$

$$LI_{RT} = [(L - R)/(L + R)] * 100$$

For  $LI_{ACC}$  calculations, R and L represent accuracy of reading words presented in the RVF and LVF, respectively, in the single-language context, in the dual-language context, or across both contexts. For  $LI_{RT}$  calculations, R and L represent response times (reading onsets) for words presented in the RVF and LVF, respectively, in the single-language context, in the dual-language context, or across both contexts. The obtained results allowed us to determine which visual half-field, and also indirectly which cerebral hemisphere, was the dominant one in the processing of Polish and English words for each participant. In the case of both  $LI_{ACC}$  and  $LI_{RT}$ , positive values indicated right visual field/left hemisphere advantage in reading words of a given language, whereas negative values—left visual field/right hemisphere advantage in the task in question.

Finally, to investigate whether or not the representations of both L1 and L2 share any common organizational features,

we performed a correlational analysis of the obtained LIs, as well as additional pairwise comparisons. Significant correlations between LIs for L1 and L2 in each of the contexts would indicate that the lateralization of the first and second language in highly-proficient bilinguals from our sample depends on one another, although they may not necessarily be similarly represented in the brains of the participants. A lack of correlations would suggest that these languages are represented independently, or even quite separately, even if they are not lateralized differently. On the other hand, significant differences obtained between LIs for L1 and L2 might indicate that one of the hemispheres is differently involved in the processing of words from each of these two languages.

All statistical analyses were carried out using SPSS 20.0 (SPSS Ins., Chicago, IL).

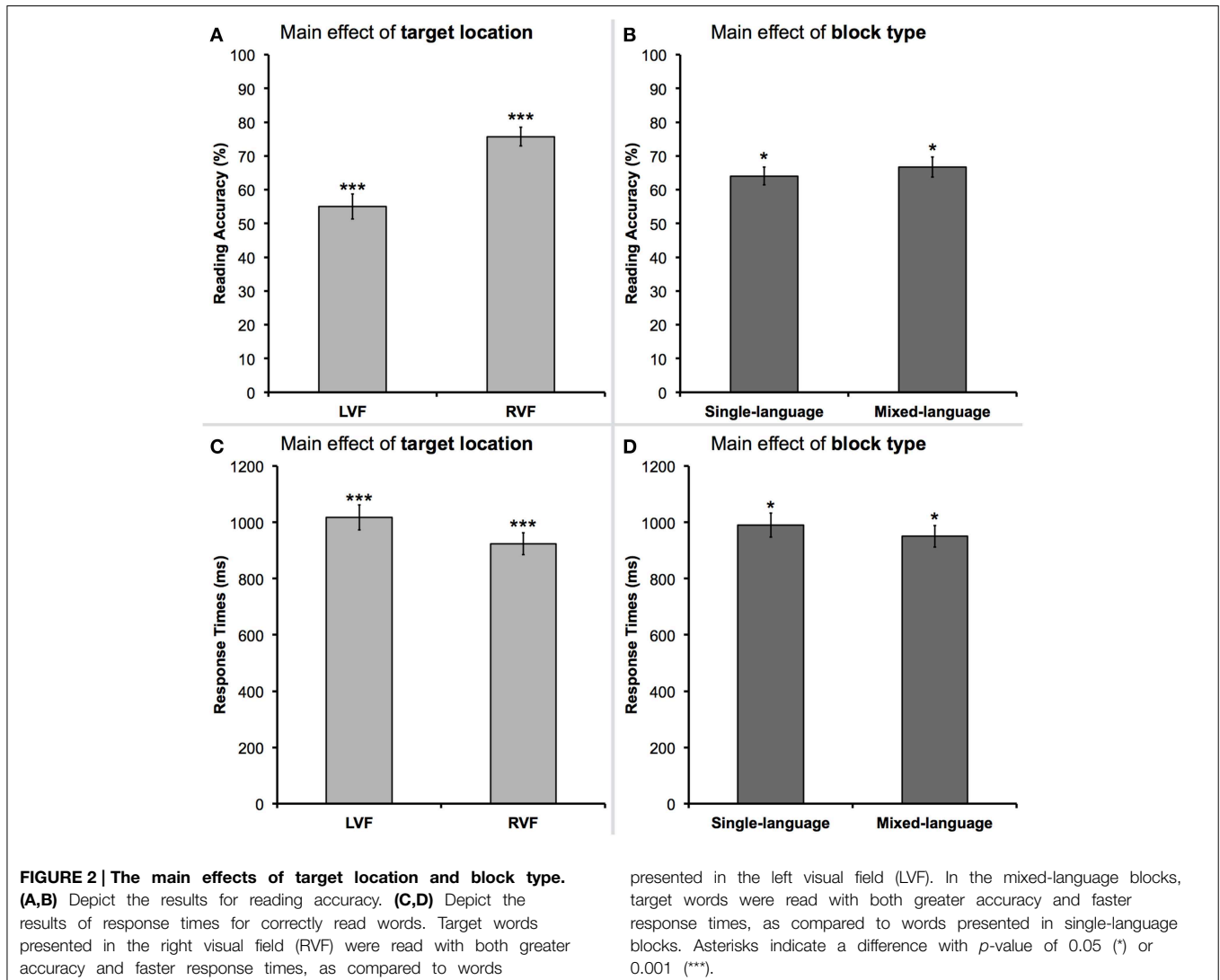
## Results

### Reading Accuracy

There was a main effect of *target location* [ $F_{(1, 17)} = 33.6$ ,  $p < 0.001$ , Partial Eta Squared ( $\rho\eta^2$ ) = 0.66], such that words presented in the RVF were read more accurately than words presented in the LVF [average reading accuracy in the RVF = 75.7%, standard error ( $SE$ ) = 2.8% vs. LVF = 55.0%,  $SE = 3.7%$ ]. This effect is shown in **Figure 2A**. We also observed a main effect of *block type* [ $F_{(1, 17)} = 4.6$ ,  $p < 0.05$ ,  $\rho\eta^2 = 0.21$ ], although quite counterintuitively the words in mixed-language blocks were read more accurately than words in single-language blocks (average accuracy of reading in mixed-language blocks = 66.7%,  $SE = 2.9%$  vs. single-language blocks = 64.0%,  $SE = 2.6%$ ). This effect is depicted in **Figure 2B**. There was also a trend toward a main effect of *target language* [ $F_{(1, 17)} = 3.1$ ,  $p = 0.10$ ,  $\rho\eta^2 = 0.16$ ]. Namely, participants tended to read target words in Polish with greater accuracy as compared to words in English (average reading accuracy in Polish = 66.5%,  $SE = 2.8%$  vs. English = 64.2%,  $SE = 2.9%$ ). No further significant effects were found, including the lack of clear trends toward interactions.

### Response Times (RTs) for Correctly Read Words

Similarly to the analysis of reading accuracy, the predicted main effect of *target location* [ $F_{(1, 17)} = 18.4$ ,  $p < 0.001$ ,  $\rho\eta^2 = 0.52$ ] was observed. Namely, for the correctly read words presented in the RVF, response times were significantly faster than for the correctly read words presented in the LVF (mean RT in the RVF = 923 ms,  $SE = 39$  ms vs. LVF = 1017 ms,  $SE = 44$  ms]. This effect is shown in **Figure 2C**. A main effect of *block type* [ $F_{(1, 17)} = 5.6$ ,  $p < 0.05$ ,  $\rho\eta^2 = 0.25$ ] revealed that participants took longer to read words in single-language blocks than in mixed-language blocks (mean RT for single-language blocks = 990 ms,  $SE = 43$  ms vs. mixed-language blocks = 951 ms,  $SE = 38$  ms). This effect is shown in **Figure 2D**. There was also a main effect of *target language* [ $F_{(1, 17)} = 11.4$ ,  $p < 0.01$ ,  $\rho\eta^2 = 0.40$ ], such that participants read words in Polish significantly faster than words in English (mean RT for Polish words = 947 ms,  $SE = 42$  ms vs. English = 994 ms,  $SE = 39$  ms). No other effects reached or even approached significance level. The mean RTs, as



well as average accuracy data, for all the conditions are listed in Table 1.

### Laterality Indices (LIs)

The results of correlational analyses are shown in Table 2. As expected, we found strong significant correlations between individuals' Polish and English LIs, for both reading accuracy and response times, in single-language context, in dual-language context, as well as across both contexts. The latter effects are shown in Figures 3A,B. Importantly, in the single-language context there was no significant difference between RT-based LIs for both languages. Individual LIs for the single-language context are shown in Figure 4A, and mean LIs in Figure 4B. In the dual-language context, however, we observed a significant right visual-field/left hemispheric advantage for reading Polish, as compared to English, words [Polish  $LI = 6.2$ ,  $SE = 1.4$  vs. English  $LI = 4.1$ ,  $SE = 1.2$ ;  $t_{(17)} = 2.4$ ,  $p < 0.05$ ]. Individual LIs for the dual-language context are shown in Figure 4C, and mean LIs, as well as a significant difference between them, in Figure 4D.

### Post-hoc Analyses and Results for the Exclusion of Possible Interpretations

To rule out the possibility that the differences between L1 and L2 reading latencies were caused by variations in voice-key sensitivity, we carried out a *post-hoc* analysis of the voicing of initial phonemes for the tested words. Voicing has been previously shown to affect the measured response times, with most voiced phonemes being detected faster than voiceless phonemes (Kessler et al., 2002). With this in mind, we ran a repeated-measures ANOVA for the frequencies (expressed in % correct) with which words from both languages were accurately read in each of the experimental conditions. The within-subjects factors were *block type* (single-language, mixed-language) and *target location* (left, right), whereas the between-subjects factors were *target language* (Polish, English) and *voicing* (voiced, voiceless). Neither the main effect of *voicing* [ $F_{(1, 76)} = 2.6$ ,  $p = 0.11$ ] nor any interactions including this factor were statistically significant.

**TABLE 1 | Block type (single-language, mixed-language), target language (Polish, English), target location (Left Visual Field, LVF; Right Visual Field, RVF) with their mean response times (ms), accuracy (%), and their standard errors of the means.**

Trial type			Response times (ms)	St. error	Accuracy (%)	St. error	N
Single-language	Polish	LVF	1002	57	58.2	3.3	18
		RVF	918	39	74.2	3.1	18
	English	LVF	1070	50	50.8	3.7	18
		RVF	972	40	72.9	3.1	18
Mixed-language	Polish	LVF	990	43	55.0	4.4	18
		RVF	877	43	78.8	3.4	18
	English	LVF	1007	39	56.0	5.0	18
		RVF	928	39	77.1	2.9	18

**TABLE 2 | The table shows the *p*-values (and *r*-values) of the correlations between the Laterality Indices (LIs) calculated for Polish and English within each of the experimental conditions (single-language, dual-language), as well as across them (general).**

	Polish, single	Polish, dual	Polish, general	English, single	English, dual	English, general
<b>CORRELATIONS FOR READING ACCURACY LIs (<math>LI_{ACC}</math>)</b>						
Polish, single	–	<b>0.18 (0.33)</b>	<0.001 (0.73)	<b>&lt;0.05 (0.53)</b>	<0.05 (0.55)	<0.01 (0.62)
Polish, dual	–	–	<0.001 (0.88)	<0.05 (0.57)	<b>&lt;0.001 (0.80)</b>	<0.001 (0.80)
Polish, general	–	–	–	<0.01 (0.68)	<0.001 (0.84)	<b>&lt;0.001 (0.88)</b>
English, single	–	–	–	–	<b>&lt;0.05 (0.52)</b>	<0.001 (0.81)
English, dual	–	–	–	–	–	<0.001 (0.92)
English, general	–	–	–	–	–	–
<b>CORRELATIONS FOR RESPONSE TIME LIs (<math>LI_{RT}</math>)</b>						
Polish, single	–	<b>0.07 (0.44)</b>	<0.001 (0.82)	<b>&lt;0.01 (0.61)</b>	0.11 (0.39)	<0.05 (0.56)
Polish, dual	–	–	<0.001 (0.88)	<0.05 (0.56)	<b>&lt;0.001 (0.78)</b>	<0.001 (0.73)
Polish, general	–	–	–	<0.01 (0.68)	<0.01 (0.70)	<b>&lt;0.001 (0.76)</b>
English, single	–	–	–	–	<b>&lt;0.01 (0.67)</b>	<0.001 (0.92)
English, dual	–	–	–	–	–	<0.001 (0.91)
English, general	–	–	–	–	–	–

The upper part of the table reports the correlations between LIs calculated on the basis of reading accuracy, whereas the lower part reports the correlations between LIs calculated on the basis of response times. Pairs of LIs that were of particular interest are highlighted in bold. Additionally, shaded cell indicate the pair wherein LIs significantly differed from each other.

Because the aforementioned analysis demonstrated that correct responses to voiced and voiceless phonemes were in fact distributed equally across different conditions therefore any differences with which they would be recorded by voice-key should not play a role. Consistent with such a hypothesis, except for the main effect of voicing [ $F_{(1, 76)} = 18.9, p < 0.001$ ] such that reading onset of words starting with voiced phonemes was indeed detected significantly faster, (and the familiar main effect of side, such that words in the right visual field were read significantly faster than words in the left visual field), none of the remaining main effects, nor One- or Two-Way interactions even approached significance level, and a trend in the Four-Way interaction was completely irrelevant to the findings reported here.

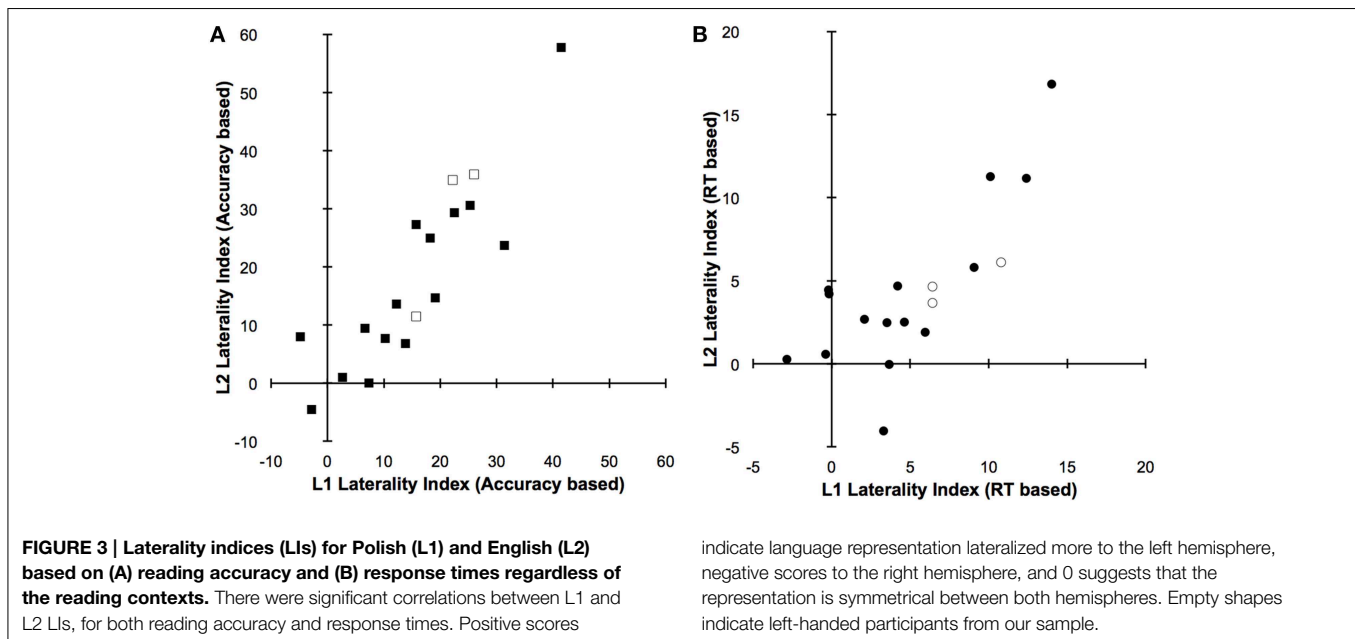
To rule out the possibility that any effects observed in the final mixed-language blocks might be due to practice effects (e.g., Garofeanu et al., 2004), we ran two 4 (*block number*)  $\times$  2 (*target location*) repeated measures ANOVAs for accuracy and response times to correctly read words in single-language blocks. There was no evidence that participants' accuracy increased with

practice in consecutive blocks, as revealed by no main effect of *block number* [ $F_{(3, 51)} = 0.5, p = 0.71, p\eta^2 = 0.03$ ]. In fact, after initial (non-significant,  $p = 0.32$ ) improvement in the second block, accuracy in the last block decreased. Moreover, there was no evidence that participants' performance, as measured by response latencies, improved with practice in consecutive blocks. This was revealed by no main effect of *block number* [ $F_{(3, 51)} = 0.6, p = 0.61, p\eta^2 = 0.04$ ]. In fact, after initial (non-significant,  $p = 0.13$ ) improvement (i.e., response time decrease) in the second block, response times increased in the subsequent blocks.

## Discussion

In this study we examined the lateralization pattern of overt word reading in single- and dual-language contexts in highly proficient Polish-English bilinguals. It was possible thanks to the utilization of the visual half-field paradigm in which in the single-language blocks only words from one language were presented and read, whereas in the mixed-language blocks words from both languages were presented and read.





Both for accuracy and for response times (or reading latencies), there was a greater advantage for reading words presented in the RVF, as opposed to the LVF. Such effects as superior accuracy of word processing and shorter response latencies that accompany a given task performed in the RVF—irrespective of the language in use—clearly indicate that the bilingual mechanisms involved in task performance both in L1 and L2 are predominantly lateralized to the left hemisphere. These results are consistent with the well-established findings that in the vast majority of people, irrespective of handedness, the number of languages acquired, and the bilingual (or even multilingual) status, language and its related skills, such as gestures, are typically represented in the left hemisphere or are at least mediated by critical left-lateralized mechanisms, including access to relevant concepts (e.g., Knecht et al., 2000; Vingerhoets et al., 2003; Króliczak et al., 2011; Vingerhoets et al., 2013; see also Króliczak, 2013; for review, see Hull and Vaid, 2007).

Despite high bilingual proficiency and the resulting lack of differences in L1 and L2 reading accuracy, the words in English were nonetheless read significantly slower than words in the native Polish. Of course, any simple differences between L1 and L2 in response times could be accounted for by the frequency of use of words from both languages in daily communication. Indeed, this interpretation is consistent with the findings that, unlike language proficiency, the daily pattern of bilingual language use is often not correlated with onset age of bilingualism (Flege et al., 2002), and may even be negatively correlated (Luk and Bialystok, 2013). In consequence, not only the activation of L2 phonology may be delayed (Spalek et al., 2014), but also the less rehearsed English may put greater motor demands on word articulation (cf. Parker Jones et al., 2011).

Counter to earlier reports suggesting that in comparison to other tongues, English is one of the most left-lateralized

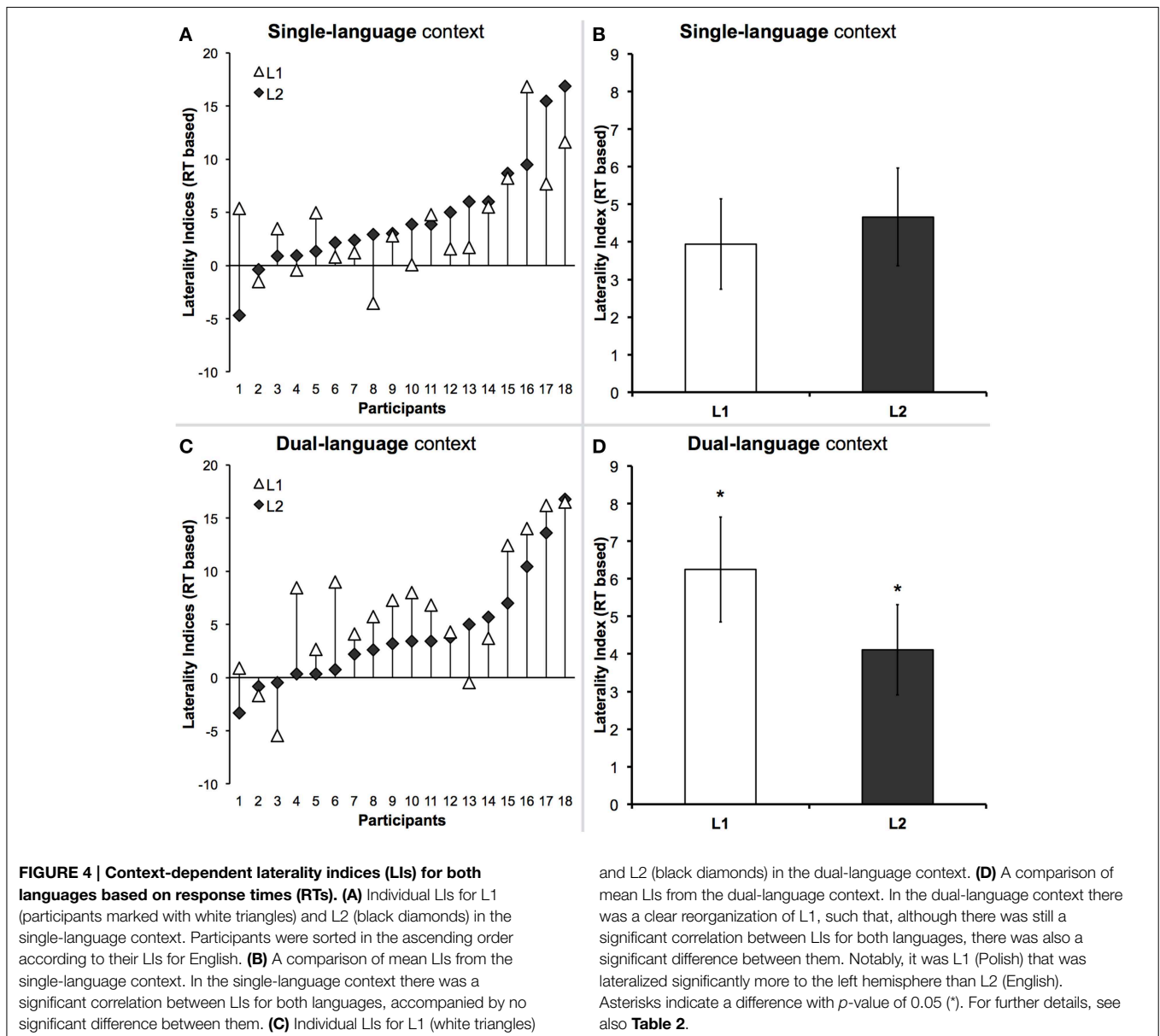
languages (e.g., Newman et al., 1999; Halsband, 2006), our results indicate that this is not always the case. Here, in the single-language context the two languages tested were similarly lateralized, whereas in the dual-language context it was the native Polish that showed greater left-sided laterality. Despite these differences, which were clearly dependent on the context, the laterality of both languages was nevertheless strongly correlated. Namely, the direction and strength of laterality for one language was always followed by a similar effect for the other, including the very rare reversed (right-sided) laterality for both. This observation is no doubt consistent with the idea that in a bilingual brain there are common mechanisms, perhaps at several different levels of language processing, that enable the fluent command of the acquired languages (e.g., Dijkstra and van Heuven, 2012).

As such, our results demonstrate that the visual half-field paradigm is not only a great method for measuring the lateralization of language, but can be equally effective in testing asymmetries of language processing in different contexts.

### The Functional Separation of L1 and L2 in the Dual-language Context

As demonstrated by no effect of language on reading accuracy, the tested group did consist of highly proficient bilinguals. To our surprise, for such individuals, reading in the single-language context was much harder than performing the same task in the dual-language context (cf. Canseco-Gonzalez et al., 2010; see also Cheng and Howard, 2008). Importantly, this effect is consistent with slower responses in the single-language context and, together, these results suggest greater within- than between-language interference, regardless of whether L1 or L2 is tested.

Although we hypothesized that the requirements for reading would increase in the dual-language context and, therefore, even if unintentionally, could lead to reliance on the same neural circuits, this was not the case. On the contrary, the most



critical outcome of this study is the observation that despite the common direction of hemispheric asymmetries, as shown by strong correlations between the LIs for the two languages, their pattern undergoes a significant functional reorganization in the dual-language context. This outcome is consistent with earlier studies showing the effects of context in which a bilingual language user operates at a given time on task performance (e.g., Marian and Spivey, 2003a,b; Canseco-Gonzalez et al., 2010). Specifically, in the paradigm used here, in the single-language context the comparison of LIs for both languages did not reveal any differences in the strength of their asymmetry. Conversely, in the dual-language context, L1-related reading mechanisms were significantly more strongly lateralized to the left hemisphere than the mechanisms for reading in L2. Indeed, this unexpected shift, typically in the form of increased left-sided L1 laterality, was

also somewhat unpredictable because when L1 performance in single- and dual-language contexts was compared, there were no significant correlations between LIs both for accuracy and for response times, whereas these correlations were still present for reading in L2.

These results strongly indicate that when two languages must be available at the same time the mechanisms involved in their control get functionally separated rather than merged. The consequence of such reorganization, either automatic or strategic, could be the minimization of the costs of maintaining readiness of the two languages and/or the increase of the efficacy of using them in parallel (cf. Christoffels et al., 2007; Cheng and Howard, 2008). This scenario—a separation of the mechanisms involved in lexical and/or phonological access—is way more likely than any reorganization of the laterality of tool-use

concepts, which in the majority of individuals should still be strongly left-lateralized (Króliczak and Frey, 2009; Michałowski and Króliczak, 2015).

## Limitations of the Study

The paradigm could benefit from the monitoring of eye movements, although the simultaneous bilateral presentation of the target and non-target words with an additional central cue controlling participants' attention should successfully prevent participants from making express saccades toward the target word when it is still visible. The immediate backward masking procedure, on the other hand, makes a regular saccade in that direction rather useless (Helon and Króliczak, 2014). Moreover, the inclusion of pseudowords as non-target stimuli could shed some new light on the possible within- and between-language interference effects observed and discussed here.

## Conclusions

All in all, this study convincingly demonstrates that the asymmetries of language processing in the bilingual brain can be effectively probed with the use of the visual half-field paradigm. Based on responses to words presented in the dominant and non-dominant visual fields, the obtained laterality indices reveal differential involvement of the co-lateralized bilingual mechanisms in such a basic linguistic task as overt reading, depending on the number of languages a proficient bilingual uses in a given context. These results clearly indicate that one of the ways of obtaining highly proficient command of two or more

languages is their functional separation at some intermediate level, whereby the lexical access is accompanied by weaker between-language interference. Thus, the adoption of a paradigm similar to the one used here opens a promising avenue for future research aimed at investigating the control mechanisms involved in the context-dependent utilization of linguistic skills in bilingual and multilingual individuals.

## Author Contributions

This project was conceptualized by MK and GK. Data was collected by MK and BM, analyzed by MK, GK, BM, and JK, and interpreted by all the authors. The manuscript was written by GK, BM, JK, and MK.

## Acknowledgments

This work is a part of a greater project supported by the Polish National Science Center (Narodowe Centrum Nauki, NCN) grant Maestro 2011/02/A/HS6/00174 to GK. During the preparation of this manuscript BM and GK were supported by the Maestro grant. The equipment used was funded by the *Ministry of Science and Higher Education* (Ministerstwo Nauki i Szkolnictwa Wyższego, MNiSW) grant 6168/IA/128/2012 to GK.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01159>

## References

- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol. (Amst.)* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014
- Bidula, S. P., and Króliczak, G. (2015). Structural asymmetry of the insula is linked to the lateralization of gesture and language. *Eur. J. Neurosci.* 41, 1438–1447. doi: 10.1111/ejn.12888
- Binkofski, F., and Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain Lang.* 127, 222–229. doi: 10.1016/j.bandl.2012.07.007
- 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
- Cheng, Y.-L., and Howard, D. (2008). The time cost of mixed-language processing: an investigation. *Int. J. Bilingualism* 12, 209–222. doi: 10.1177/1367006908098568
- 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
- Cohen, L., and Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage* 22, 466–476. doi: 10.1016/j.neuroimage.2003.12.049
- Dijkstra, T., and van Heuven, W. J. B. (2012). “Word recognition in the bilingual brain,” in *The Handbook of the Neuropsychology of Language*, Vols. 1, 2, ed M. Faust (Oxford: Wiley-Blackwell), 451–471. doi: 10.1002/9781118432501.ch22
- Flege, J. E., Mackay, I. R. A., and Piske, T. (2002). Assessing bilingual dominance. *Appl. Psycholinguist.* 23, 567–598. doi: 10.1017/S0142716402004046
- Garofeanu, C., Króliczak, G., Goodale, M. A., and Humphrey, G. K. (2004). Naming and grasping common objects: a priming study. *Exp. Brain Res.* 159, 55–64. doi: 10.1007/s00221-004-1932-z
- Grosjean, F. (2001). “The bilingual’s language modes,” in *One Mind, Two Languages: Bilingual Language Processing*, ed L. N. Janet (Oxford: Blackwell), 1–22.
- Grossi, G., Savill, N., Thomas, E., and Thierry, G. (2010). Posterior N1 asymmetry to English and Welsh words in Early and Late English–Welsh bilinguals. *Biol. Psychol.* 85, 124–133. doi: 10.1016/j.biopsycho.2010.06.003
- Halsband, U. (2006). Bilingual and multilingual language processing. *J. Physiol. Paris* 99, 355–369. doi: 10.1016/j.jphysparis.2006.03.016
- Helon, H., and Króliczak, G. (2014). The effects of visual half-field priming on the categorization of familiar intransitive gestures, tool use pantomimes, and meaningless hand movements. *Front. Psychol.* 5:454. doi: 10.3389/fpsyg.2014.00454
- Hull, R., and Vaid, J. (2007). Bilingual language lateralization: a meta-analytic tale of two hemispheres. *Neuropsychologia* 45, 1987–2008. doi: 10.1016/j.neuropsychologia.2007.03.002
- Hunter, Z. R., and Brysbaert, M. (2008). Visual half-field experiments are a good measure of cerebral language dominance if used properly: evidence from fMRI. *Neuropsychologia* 46, 316–325. doi: 10.1016/j.neuropsychologia.2007.07.007
- Kessler, B., Treiman, R., and Mullenix, J. (2002). Phonetic biases in voice key response time measurements. *J. Mem. Lang.* 47, 145–171. doi: 10.1006/jmla.2001.2835
- Klein, D., Zatorre, R. J., Chen, J.-K., Milner, B., Crane, J., Belin, P., et al. (2006). Bilingual brain organization: a functional magnetic resonance adaptation study. *Neuroimage* 31, 366–375. doi: 10.1016/j.neuroimage.2005.12.012

- Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., et al. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain* 123, 2512–2518. doi: 10.1093/brain/123.12.2512
- Króliczak, G. (2013). Praxis in left-handers. *Kultura i Edukacja (Cult. Educ.)* 6, 5–31.
- Króliczak, G., and Frey, S. H. (2009). A common network in the left cerebral hemisphere represents planning of tool use pantomimes and familiar intransitive gestures at the hand-independent level. *Cereb. Cortex* 19, 2396–2410. doi: 10.1093/cercor/bhn261
- Króliczak, G., Piper, B. J., and Frey, S. H. (2011). Atypical lateralization of language predicts cerebral asymmetries in parietal gesture representations. *Neuropsychologia* 49, 1698–1702. doi: 10.1016/j.neuropsychologia.2011.02.044
- Leonard, M. K., Brown, T. T., Travis, K. E., Gharapetian, L., Hagler, D. J., Dale, A. M., et al. (2010). Spatiotemporal dynamics of bilingual word processing. *Neuroimage* 49, 3286–3294. doi: 10.1016/j.neuroimage.2009.12.009
- 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
- Marian, V., and Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Appl. Psycholinguist.* 24, 173–193. doi: 10.1017/S0142716403000092
- Marian, V., and Spivey, M. (2003b). Competing activation in bilingual language processing: within- and between-language competition. *Bilingualism Lang. Cogn.* 6, 97–115. doi: 10.1017/S1366728903001068
- McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299. doi: 10.1016/S1364-6613(03)00134-7
- McNair, N. A., and Harris, I. M. (2012). Disentangling the contributions of grasp and action representations in the recognition of manipulable objects. *Exp. Brain Res.* 220, 71–77. doi: 10.1007/s00221-012-3116-6
- Meschyan, G., and Hernandez, A. E. (2006). Impact of language proficiency and orthographic transparency on bilingual word reading: an fMRI investigation. *Neuroimage* 29, 1135–1140. doi: 10.1016/j.neuroimage.2005.08.055
- Michałowski, B., and Króliczak, G. (2015). Sinistrals are rarely “right”: evidence from tool-affordance processing in visual half-field paradigms. *Front. Hum. Neurosci.* 9:166. doi: 10.3389/fnhum.2015.00166
- Newman, A., Izvorski, R., Davis, L., Neville, H., and Ullman, M. T. (1999). Distinct electrophysiological patterns in the processing of regular and irregular verbs. *J. Cogn. Neurosci.* (Suppl. 47).
- Parker Jones, Ö., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., et al. (2011). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cereb. Cortex* 22, 892–902. doi: 10.1093/cercor/bhr161
- Peng, G., and Wang, W. S.-Y. (2011). Hemisphere lateralization is influenced by bilingual status and composition of words. *Neuropsychologia* 49, 1981–1986. doi: 10.1016/j.neuropsychologia.2011.03.027
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain* 121, 1841–1852. doi: 10.1093/brain/121.10.1841
- Spalek, K., Hoshino, N., Wu, Y. J., Damian, M., and Thierry, G. (2014). Speaking two languages at once: unconscious native word form access in second language production. *Cognition* 133, 226–231. doi: 10.1016/j.cognition.2014.06.016
- Van Heuven, W. J. B., and Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Res. Rev.* 64, 104–122. doi: 10.1016/j.brainresrev.2010.03.002
- Vingerhoets, G., Alderweireldt, A.-S., Vandemaële, P., Cai, Q., Van der Haegen, L., Brysbaert, M., et al. (2013). Praxis and language are linked: evidence from co-lateralization in individuals with atypical language dominance. *Cortex* 49, 172–183. doi: 10.1016/j.cortex.2011.11.003
- Vingerhoets, G., Van Borsel, J., Tesink, C., van den Noort, M., Deblaere, K., Seurinck, R., et al. (2003). Multilingualism: an fMRI study. *NeuroImage* 20, 2181–2196. doi: 10.1016/j.neuroimage.2003.07.029
- Wu, Y. J., and Thierry, G. (2010). Investigating bilingual processing: the neglected role of language processing contexts. *Front. Psychol.* 1:178. doi: 10.3389/fpsyg.2010.00178

**Conflict of Interest Statement:** 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 © 2015 Krefta, Michałowski, Kowalczyk and Króliczak. 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) or licensor 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.

