

# LANGUAGE ACQUISITION IN DIVERSE LINGUISTIC, SOCIAL AND COGNITIVE CIRCUMSTANCES

EDITED BY: Maria Garraffa, Maria Teresa Guasti, Theodoros Marinis and Gary Morgan

PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2019 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 978-2-88945-689-5

DOI 10.3389/978-2-88945-689-5

## About Frontiers

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

## Frontiers Journal Series

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

## Dedication to Quality

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

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

## What are Frontiers Research Topics?

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

# LANGUAGE ACQUISITION IN DIVERSE LINGUISTIC, SOCIAL AND COGNITIVE CIRCUMSTANCES

Topic Editors:

**Maria Garraffa**, Heriot-Watt University, United Kingdom

**Maria Teresa Guasti**, Università Milano Bicocca, Italy

**Theodoros Marinis**, University of Konstanz, Germany

**Gary Morgan**, City University of London, United Kingdom



Language building blocks.

Image: Iacopo Leardini.

The language experience of children developing in linguistically diverse environments is subject to considerable variation both in terms of quantity and quality of language exposure. It is an open question how to investigate language exposure patterns and more important which factors are relevant for successful language learning.

For example, children acquiring a minority language, including a signed language, are exposed to less variety of input than children acquiring a more global language. This is because they are living in a smaller linguistic community and with fewer occasions to use the language in everyday life. Despite this reduced input, most native signers are successful language learners. In contrast native language competence is not always achieved in signing deaf children with hearing parents or those with cochlear implants learning a spoken language. A similar outcome but with very different reasons has also been reported for hearing children with language impairment. In these populations acquisition of morphosyntactic aspects is developing atypically ending with an uncomplete linguistic repertoire.

The circumstances of exposure during language development tend to differ in significant ways with respect to a large number of factors, such as, (i) length, quality and quantity of input, (ii) social status and attitudes toward the language, (iii) cognitive abilities required for language learning, and (iv) age of first exposure. Having early exposure to a range of different speakers is important in the acquisition of any language and may affect language proficiency. However, negative societal attitudes or a cognitive based disadvantage may create an unfavourable learning environment that prevents language learning from surfacing typically. This situation inevitably generates a different type of exposure for the child and consequently different language competence.

In this Research Topic we intend to encourage the debate on social, linguistic and cognitive factors at play for designing an effective environment for language acquisition aiming at integrating linguistic variables coming from theoretical studies on language with environmental variables, such as, measures of language input or cognitive abilities on functions ancillary to language development.

**Citation:** Garraffa, M., Guasti, M. T., Marinis, T., Morgan, G., eds. (2019). Language Acquisition in Diverse Linguistic, Social and Cognitive Circumstances. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-689-5

# Table of Contents

**06 Editorial: Language Acquisition in Diverse Linguistic, Social and Cognitive Circumstances**

Maria Garraffa, Maria Teresa Guasti, Theodoros Marinis and Gary Morgan

## **SECTION 1**

### **LANGUAGE ACQUISITION AT DIFFERENT DEVELOPMENTAL STAGES**

**09 Internal Grammar and Children's Grammatical Creativity Against Poor Inputs**

Adriana Belletti

**16 The Peaceful Co-existence of Input Frequency and Structural Intervention Effects on the Comprehension of Complex Sentences in German-Speaking Children**

Flavia Adani, Maja Stegenwallner-Schütz and Talea Niesel

**27 Comprehension of Subject and Object Relative Clauses in a Trilingual Acquisition Context**

Angel Chan, Si Chen, Stephen Matthews and Virginia Yip

**44 Home Language Will not Take Care of Itself: Vocabulary Knowledge in Trilingual Children in the United Kingdom**

Karolina Mieszkowska, Magdalena Łuniewska, Joanna Kotak, Agnieszka Kacprzak, Zofia Wodniecka and Ewa Haman

**55 Pronoun Interpretation in the Second Language: Effects of Computational Complexity**

Roumyana Slabakova, Lydia White and Natália Brambatti Guzzo

**67 Production is Only Half the Story — First Words in Two East African Languages**

Katherine J. Alcock

**81 When Meaning is not Enough: Distributional and Semantic Cues to Word Categorization in Child Directed Speech**

Sara Feijoo, Carmen Muñoz, Anna Amadó and Elisabet Serrat

## **SECTION 2**

### **MULTILINGUAL CHILDREN AND CHILDREN WITH LANGUAGE DISORDERS: SIMILARITIES AND DIFFERENCES**

**92 Language Proficiency and Sustained Attention in Monolingual and Bilingual Children With and Without Language Impairment**

Tessel Boerma, Paul Leseman, Frank Wijnen and Elma Blom

**108 Methods for Identifying Specific Language Impairment in Bilingual Populations in Germany**

Cornelia Hamann and Lina Abed Ibrahim

### SECTION 3

#### LANGUAGE ACQUISITION WITH MINORITY LANGUAGES

- 127** *Verbal Working Memory is Related to the Acquisition of Cross-Linguistic Phonological Regularities*  
Evelyn Bosma, Wilbert Heeringa, Eric Hoekstra, Arjen Versloot and Elma Blom
- 138** *How Does L1 and L2 Exposure Impact L1 Performance in Bilingual Children? Evidence From Polish-English Migrants to the United Kingdom*  
Ewa Haman, Zofia Wodniecka, Marta Marecka, Jakub Szewczyk, Marta Białecka-Pikul, Agnieszka Otwinowska, Karolina Mieszkowska, Magdalena Łuniewska, Joanna Kołak, Aneta Mięgisz, Agnieszka Kacprzak, Natalia Banasik and Małgorzata Foryś-Nogala
- 159** *Acquiring Clitic Placement in Bilectal Settings: Interactions Between Social Factors*  
Kleanthes K. Grohmann, Elena Papadopoulou and Charalambos Themistocleous
- 173** *The Influence of Bilectalism and Non-standardization on the Perception of Native Grammatical Variants*  
Evelina Leivada, Elena Papadopoulou, Maria Kambanaros and Kleanthes K. Grohmann
- 184** *Acquisition of Classifier Constructions in HKSL by Bimodal Bilingual Deaf Children of Hearing Parents*  
Gladys W. L. Tang and Jia Li

### SECTION 4

#### LEARNING READING AND SPEAKING IN MORE THAN ONE LANGUAGE

- 206** *Is Morphological Awareness a Relevant Predictor of Reading Fluency and Comprehension? New Evidence From Italian Monolingual and Arabic-Italian Bilingual Children*  
Mirta Vernice and Elena Pagliarini
- 221** *Assessing the Formation of Experience-Based Gender Expectations in an Implicit Learning Scenario*  
Anton Öttl and Dawn M. Behne



# Editorial: Language Acquisition in Diverse Linguistic, Social and Cognitive Circumstances

Maria Garraffa<sup>1\*</sup>, Maria Teresa Guasti<sup>2</sup>, Theodoros Marinis<sup>3,4</sup> and Gary Morgan<sup>5</sup>

<sup>1</sup> Psychology Department, Heriot-Watt University, Edinburgh, United Kingdom, <sup>2</sup> Psychology Department, Università Milano Bicocca, Milan, Italy, <sup>3</sup> Linguistics Department, University of Konstanz, Konstanz, Germany, <sup>4</sup> Department of Clinical Language Sciences, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom, <sup>5</sup> Division of Language & Communication Science, City University London, London, United Kingdom

**Keywords:** language acquisition, language pathology, multilingualism, heritage language (HL), second language (L2) acquisition, sign language (SL), syntax, vocabulary acquisition

## Editorial on the Research Topic

### Language Acquisition in Diverse Linguistic, Social and Cognitive Circumstances

The language experience of children growing up in linguistically diverse environments is subject to considerable variation both in terms of input quantity and quality and these factors are predictive of future language abilities (e.g. Hart and Risley, 1995). While virtually all typically developing (TD) children acquire language competence, there are large differences in the extent to which vocabulary and higher-level linguistic skills develop, especially in children with atypical language development. This research topic encouraged a debate around the linguistic and environmental factors at play in a set of diverse environments for language acquisition. Language acquisition cannot be investigated without a clear description of the linguistic phenomena that need to be acquired. It is not clear, for example, why some phenomena are acquired later and some earlier; and if differences between children in processing are an effect of differences in competence, or differences in levels of cognitive variables such as non-verbal IQ, working memory, or Executive Functions.

A first theme emerging from the contributions in this research topic is the different trajectories of linguistic phenomena at different developmental stages. Finer aspects of language acquisition do not come from the environment but from maturational changes in early learners. This is the case in Belletti and a study of the children's ability to answer direct object questions. Productions reported by Italian children are non-attested in adults' grammar but are compatible with an immature grammatical system. The study supports the idea that input is not a sufficient variable to explain development and also the outcomes of the study are compatible with developmental trajectories.

A further step in the debate on how to integrate environmental and internal (biological) factors was discussed in a study on German preschool children's comprehension of Relative Clauses (RC). Age modulated the comprehension of Object RCs, with older children being more sensitive to pure grammatical distinctions compared to younger children who were more affected by non-linguistic cues (Adani et al.).

The comprehension of RCs in a trilingual group of children with Cantonese (L1), Mandarin (L2), and English (L3) was investigated by Chan et al. that looked at the effect of limited exposure due to the multilingual environment. Transfer from the head-initial language (English) was reported in the trilingual group in the comprehension of object RCs in Cantonese because of structural overlap and intensive exposure to English. The study points out the importance of identification of vulnerable domains, such as Head noun assignment in object RCs in multilingual Cantonese children acquiring English.

Another group of trilingual children with developmental vulnerability due to scarce input was investigated in a study of vocabulary skills, comparing monolingual, bilingual, and trilingual

## OPEN ACCESS

### Edited and reviewed by:

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

### \*Correspondence:

Maria Garraffa  
m.garraffa@hw.ac.uk

### Specialty section:

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

**Received:** 03 August 2018

**Accepted:** 05 September 2018

**Published:** 24 October 2018

### Citation:

Garraffa M, Guasti MT, Marinis T and  
Morgan G (2018) Editorial: Language  
Acquisition in Diverse Linguistic,  
Social and Cognitive Circumstances.  
Front. Psychol. 9:1807.  
doi: 10.3389/fpsyg.2018.01807

children (Mieszkowska et al.). For the majority language (English) no difference was found across the three groups. However the minority language was reported as incrementally weaker in both bilingual (reduced expressive vocabulary) and trilingual (reduced expressive and receptive vocabulary) children. The authors suggest that the home language needs to be supported more to achieve a developmental trajectory consistent with the dominant language of the environment.

A well-established pattern in TD children is the greater difficulty in interpreting sentences with pronouns (in particular referential antecedents compared to quantified and full vs. reduced pronouns). Few studies have investigated the interpretation of pronouns in L2 learners. A study on adult L2 learners found that beginners' performance is affected by type of pronoun and antecedent. These results are in line with the grammar of monolingual children, advocating for a general linguistics principle at play in L2 learners (Slabakova et al.).

The authors argue that studies of the developmental trajectory of language development should include the acquisition of different word categories. A significant difference between comprehension and production of both nouns and verbs was reported in a study on child learners of two East African Languages. While the findings were in keeping with previous noun-bias work, making the study cross linguistically valid, a quantitative and qualitative difference was reported. The proportion of spoken verbs correlated with increases in vocabulary size, and with more nouns in the first spoken words and verbs in the comprehended ones (Alcock).

Feijoo et al. questioned the fundamental assumption of semantic bootstrapping in the acquisition of language categories. Investigating child-directed speech input to children under the age of 2;6 they showed that semantic cues alone are not sufficient for word categorization. Rather children need to carry out an analysis of both distributional and semantic cues in the child-directed speech. These results are in line with theories that suggest the need for an integration of multiple cues from different sources in language development.

A second theme was the challenge of how to differentiate multilingual children with slower early language development from multilingual children with developmental language disorder (DLD). While language difficulties in the two populations can look similar, further investigations have reported divergent behaviors between multilingual and atypical populations (Armon-Lotem, 2017). In a study of Dutch children's cognitive and linguistic abilities, Boerma et al. reported that auditory sustained attention mediated the effect of L1 on vocabulary and morphology in both the monolingual and multilingual groups. The study supports the idea that a weak linguistic ability in children with a developmental language disorder (DLD) can be related to an impairment in sustaining attention to auditory stimuli.

Another study on effective tools for differentiating multilingual children and children with DLD used non-word- and sentence repetition as clinical markers (Hamann and Ibrahim). The study showed that the two measures are reliable tools for identification of DLD in multilingual contexts if background information is included. Crucially, both tasks can discriminate multilingual TD children from

monolingual children with DLD and multilingual TD children from multilingual children with DLD, with sentence repetition being more affected by language dominance. The study also highlighted that testing in the home language in a heritage context might lead to unreliable classifications.

A further theme was the acquisition of minority languages, including signed languages. Bosma et al. explored the cognitive components ancillary for language acquisition, focusing on the role of verbal working memory (vWM) for the acquisition of phonological regularities in a longitudinal study in a group of Frisian-Dutch bilingual children. The study strongly supported the hypothesis that vWM is an essential component to detect phonological regularities in a task targeting cognates in the two languages.

The role of exposure was addressed in an extensive study (from single words to narratives) of bilingual Polish-English children, focusing on L1 exposure (Haman et al.). The bilingual children scored lower compared to monolinguals in all language domains except discourse, with more pronounced differences in production. Grammar scores were not related to the levels of L1, but were predicted by general cognitive abilities. L2 exposure negatively influenced productive grammar in the L1, suggesting possible L2 transfer effects. Importantly, the authors did not find any evidence that the gap between monolinguals and multilinguals would be fully closed by manipulating L1 input.

Factors affecting children acquiring a minority language should be investigated in interaction with the sociolinguistic context of acquisition. This is the case in a large-scale study on the acquisition of clitic placement in bilingual children (Grohmann et al.). The study revealed early discrimination of enclisis in Cypriot Greek and proclisis in standard Greek, but effects related to the context of acquisition, with proclisis increasing as children enter primary school, advocating for the role of formal education in bilingual settings.

A second study on bilingualism focused on speaker's perception of the two varieties, investigating the hypothesis of a grammatical fluidity in bilingual speakers (Leivada et al.). A variety-judgment task was developed in a large study on monolinguals, bilinguals, and bilinguals, including heritage language learners and L1 attriters. The study supported the idea of a different grammatical appreciation in speakers of non-standardized languages (Leivada et al.).

The role of duration of exposure was tested in a study on Deaf children immersed in a dual language input environment (Cantonese and Hong Kong Sign Language, HKSL). The study focused on the acquisition of classifier constructions in HKSL, a structure that emerges later and with cross-linguistics differences between the two languages, in particular verb root and word order. The findings revealed Deaf children's gradual convergence on the adult grammar despite late exposure to HKSL. Evidence of cross-linguistic influence on word order came from the initial adoption of a Cantonese structure. There was also a prolonged period of adherence to the SVO order across all ages (Tang and Li).

Early L2 learners revealed a different performance in reading compared to monolingual children. Vernice and Pagliarini looked at the contribution of morphological awareness to reading in a group of Italian L1 and Arabic-Italian early L2 children.

Accuracy in the morphological awareness tasks was a significant predictor of reading fluency. The study highlights the critical role of morphological processing in reading efficiency and suggests that morphological awareness training could improve reading in bilingual students.

Another contribution pointed out the role of the learning scenario in language acquisition, comparing implicit and explicit learning. To assess whether the formation of experience-based expectations is dependent on explicit awareness, Ottl and Behen presented data from an experiment in which gender coding was acquired implicitly. Results showed that participants develop frequency-based expectations comparable to those previously observed in an explicit learning scenario. At the same time, however, the study suggests that expectations surface earlier in the implicit learning scenario.

## REFERENCES

- Armon-Lotem, S. (2017). Disentangling bilingualism from SLI: dissociating exposure and input. *Bilingual. Lang. Cogn.* 20, 33–34. doi: 10.1017/S1366728916000377
- Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing Company.

## AUTHOR CONTRIBUTIONS

MG, MTG, TM, and GM drafted the work and revised it critically for important intellectual content. They did the final approval of the version to be published and are 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.

## ACKNOWLEDGMENTS

We would like to thank all 59 authors and 34 reviewers who offered their manuscripts and their constructive comments for this Research Topic.

**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 © 2018 Garraffa, Guasti, Marinis and Morgan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Internal Grammar and Children's Grammatical Creativity against Poor Inputs

Adriana Belletti<sup>1,2\*</sup>

<sup>1</sup> Département de Linguistique, Université de Genève, Geneva, Switzerland, <sup>2</sup> DISPOC, Università di Siena, Siena, Italy

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Rosalind Jean Thornton,  
Macquarie University, Australia  
David W. Lightfoot,  
Georgetown University, United States

### \*Correspondence:

Adriana Belletti  
adriana.belletti@unige.ch

### Specialty section:

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

**Received:** 28 June 2017

**Accepted:** 14 November 2017

**Published:** 01 December 2017

### Citation:

Belletti A (2017) Internal Grammar and  
Children's Grammatical Creativity  
against Poor Inputs.  
Front. Psychol. 8:2074.  
doi: 10.3389/fpsyg.2017.02074

This article is about the unexpected linguistic behavior that young children sometimes display by producing structures that are only marginally present in the adult language in a constrained way, and that adults do not adopt in the same experimental conditions. It is argued here that children's capacity to overextend the use of given syntactic structures thereby resulting in a grammatical creative behavior is the sign of an internal grammatical pressure which manifests itself given appropriate discourse conditions and factors of grammatical complexity and which does not necessarily require a rich input to be put into work. This poverty of the stimulus type situation is illustrated here through the overextended use of *a*-Topics and reflexive-causative passives by young Italian speaking children when answering eliciting questions concerning the direct object of the clause.

**Keywords:** grammatical creativity, grammatical complexity, featural relativized minimality, *a*-Topic, passive

## INTRODUCTION

Young children sometimes display an unexpected linguistic behavior: they produce structures that are at most only marginally present in the adult language. This holds both in the sense that the relevant constructions are rarely present in the language and in the sense that their occurrence is severely constrained, as descriptive work clearly indicates. Furthermore, children may react differently from adults to the very same experimental conditions, producing structures that are not "simpler" in any intuitive sense of the term. This type of children's linguistic behavior, which is in fact quite widespread in work on development, indicates that some internal pressure, partly due to factors of computational complexity as we will argue, leads children to be grammatically creative.

The following article is about two case studies with precisely these characteristics, based on experimental results on the acquisition of Italian, recently presented and discussed in detail in Belletti and Manetti (2017). The experiment aimed at studying the acquisition of two different empirical domains in Italian: Romance-type topicalization/Clitic Left Dislocation (CLLD) and types of passive. The first domain is part of a new line of research on the acquisition of different discourse-related positions in the left periphery of the clause identified by cartographic studies (Rizzi and Bocci, 2017). Specifically, the research aimed at studying the acquisition of Topic positions hosting a left dislocated direct object, which in Italian/Romance yields a so called Clitic Left Dislocation/CLLD as in *Il cane, il gatto lo lava*/the dog, the cat washes it-Cl. In CLLD the sentence following the left dislocated direct object, which is a discourse topic given in previous discourse, predicates a property of the proposed noun phrase and (obligatorily) contains a clitic pronoun referring to it (*lo* in the

example above). The second domain is a classical research topic in language acquisition, which has recently received renewed attention in the theoretical debate (see, a.o. Manetti, 2013; Snyder and Hyams, 2015; less recently Crain, 1991), i.e. the acquisition of passive, in the aim of determining which types of passive children may prefer or access first, if there is any preference or earlier access at all.

In what follows, I will briefly outline the essential relevant features of the experimental design and illustrate and discuss aspects of the results that are relevant for the present discussion. For a thorough description (of both method and materials) and overall discussion of the articulated results the reader is referred to Belletti and Manetti (2017). Overall, the results to be reviewed here provide new evidence from empirical domains that have not been previously discussed in this connection, that children's linguistic behavior does not simply mirror adult production and does not simply reflect what children hear most. In this sense then, children are capable to express an intriguing grammatical creativity that does not conform to their input (pace Tomasello, 2003, and subsequent related literature). Such creativity in turn is not unconstrained, but, as will be illustrated here, follows the principled organization of the UG-constrained internal grammatical system, thus also indicating clear continuity in the process of linguistic development.

## OUTLINE OF THE DESIGN AND OF MAIN RELEVANT RESULTS

In Belletti and Manetti's (2017) design (their Experiment 1) young children (39, age range 4;1–5;11) had to answer a question concerning the object of a transitive action. As mentioned in the introduction, the aim of the elicitation experiment was to check whether young children at the ages investigated can access left peripheral topic positions; a related aim was to also determine whether they can access passive structures and, in case they do, which type of passive they prefer among the different ones available in Italian, if there is any preference at all. The question came at the end of a short introductory story, which was accompanied by illustrating images. For instance, given a story ending with a picture showing "a giraffe licking a cow and a rabbit touching a penguin," in a Two-topic condition (i.e., a contrastive topic situation, Benincà and Poletto, 2004; Bocci, 2013), the question in (1) was asked to the children:

- (1) Che cosa succede ai miei amici, il pinguino e la mucca?  
What is happening to my friends, the penguin and the cow?

Italian speaking children (both age groups) often answered to this type of question with a CLD structure (25% of their answers). Use of a CLD structure in this discourse condition is perfectly adequate and appropriate. However, children realized the CLD in a peculiar way: Children's preposed direct object was typically introduced by preposition *a*, as illustrated in (2):

- (2) *Il coniglio a i' pinguino lo tocca*  
The rabbit to the penguin him.CL touches  
"The rabbit touches the penguin" (Adele 4;9)

I will refer to this type of preposed direct object topic as an *a*-Topic. Crucially, all children were monolingual speakers of a central (Tuscan) variety of Italian. In this variety, which corresponds to the standard one, direct objects are not introduced by preposition *a*.

Another type of answer produced by children in some cases (11% of their answers) is a passive sentence. This type of answer is also perfectly adequate and appropriate in the discourse condition created by the experiment; in fact, this is the type of answer most widely adopted by the 24 adult controls (68% of their answers), in exactly the same elicitation experiment. The passive utilized by children, however, is different from the one most typically produced by adults. Children exclusively resorted to a type of passive that is rarely present in adult Italian, a reflexive-causative passive illustrated in (3) (*si*-causative passive henceforth):

- (3) *La mucca si fa leccare dalla giraffa*  
The cow SI-makes lick by the giraffe  
"The cow makes the giraffe lick it" (Olmo, 4;1)

And indeed also in the experimental setting of the experiment, the most widely adopted type of passive in adults' answers was not a *si*-causative passive (19% of the produced passives) but rather a periphrastic passive using *essere*/be or (mostly) *venire*/come as the passive auxiliary (49% of the produced passives), as in (4) (copular/*venire* passive, henceforth):

- (4) *La mucca è/viene leccata dalla giraffa*  
the cow is/comes leaked by the giraffe  
"The cow is being licked by the giraffe"

In the following sections I discuss and motivate in some detail the relevance of these results for the issue raised in the introduction concerning children's grammatical creativity.

## THE CASE OF *a*-TOPICS IN CHILDREN'S CLLDS

In standard Italian direct objects are not introduced by a preposition, no matter what their nature is (e.g., specific or indefinite). Standard Italian is not a so-called Differential Object Marking/DOM language. Only in few cases, and marginally so for many speakers, can direct objects be realized as *a*-Topics: when they are the object experiencer of psych-verbs of the *worry* class. See the following contrast in (5), from Belletti and Rizzi (1988), quoting Benincà's observation (Benincà, 1986).

- (5) a (?) A Gianni, questi argomenti non l'hanno convinto  
to Gianni, these arguments him-CL have not convinced  
"These arguments have not convinced Gianni"
- b \*A Gianni, la gente non lo conosce  
to Gianni, people him-CL do not know  
"People do not know Gianni"

The contrast between (5)a, marginally acceptable for some speakers, and (5)b completely excluded by all Italian speakers illustrates the fact that only an object experiencer can be

(marginally) realized as an *a*-Topic. No contrast is present in (6) in which the object fills the object position and is not preposed; *a*-marking is excluded in both cases:

- (6) a \*Questi argomenti non hanno convinto a Gianni  
these arguments have not convinced to Gianni
- b \*La gente non conosce a Gianni  
people do not know to Gianni

The examples in (7) illustrate the other context in which *a*-Topics are possible in standard Italian: when the topic is a (mainly 1st or 2nd person) pronoun, possibly also allowing, in these cases, direct objects that are not experiencers (see also Renzi, 1988; Berretta, 1989 for relevant discussion):

- (7) a A te, non ti conosco  
to you I do not you-CL know  
“I do not know you”
- b A me, nessuno mi ha chiamato  
to me nobody me-CL has called  
“Nobody called me”
- c ?A lui, lo rispettano  
to him/him they him-CL respect  
“They respect him”

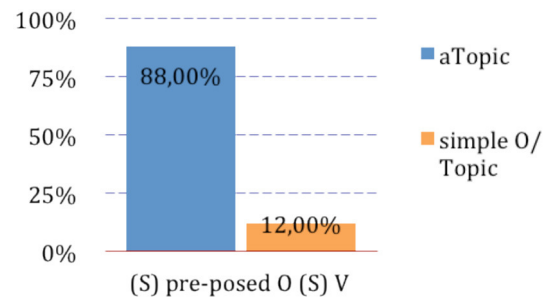
These are the two main distributional properties of *a*-Topics in standard Italian. A detailed discussion and description of the constrained distribution of *a*-Topics in Italian is provided in Belletti (2017a) where the hypothesis is put forth that *a*-Topics may be the realization of a property of the left periphery whereby the preposed object is interpreted as affected by the event described by the verb and the speaker feels particularly involved and adopts an empathic point of view towards it<sup>1</sup>. Thus, by expressing the preposed object in the form of an *a*-Topic, children have overextended the constrained and limited option of adult standard Italian in at least two ways:

- All of their *a*-Topics were lexical noun phrases (i.e., they were not—1st or 2nd person—pronouns)
- All of their *a*-Topics were objects of agentive verbs (i.e. they were not object experiencers of psych-verbs)

<sup>1</sup>A comparison with Spanish is also provided in Belletti (2017a), mainly based on Leonetti (2004) and references cited therein. A discussion of the comparative issue goes beyond the scope of the present article. It can just be noted in passing here that southern varieties of Italian are DOM languages along similar lines as Spanish. Hence, for speakers of these varieties *a*-Topics would be the reflex of DOM. A totally different situation from the one described in the text, as children were all speakers of the standard non-DOM variety of Italian as they never marked the direct object in its canonical direct object clause internal position with preposition *a*. According to Escandell-Vidal (2009) in Balearic Catalan direct objects are *a*-marked when they are (left peripheral) topics, but never when they fill the canonical direct object position, thus displaying the same syntax as the one of the Italian speaking developing children discussed in the text. This is an interesting convergence with the results from the Italian-speaking children, which also further illustrates the UG-constrained continuity of language development mentioned in the introduction. This comparative aspect of the significance of the results discussed here is further developed in current work.

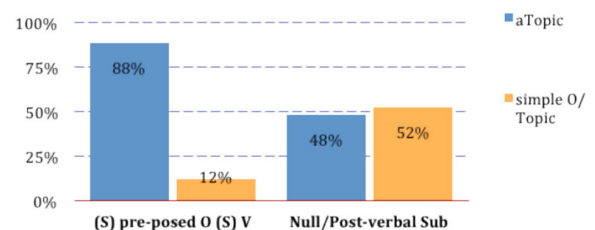
In children's experimental results from Belletti and Manetti (2017, Experiment 1), direct object topics have been realized as *a*-Topics in the vast majority of cases. Specifically, when the preverbal subject of the following clause was an overt lexical noun phrase, the topic was realized as an *a*-Topic in 88% of the cases<sup>2</sup>; it was realized as a simple direct object topic instead (with no preposition) in the remaining 12% of the cases. This is illustrated in (8). Recall that in standard Italian the latter option in which the preposed object topic is not introduced by preposition *a* (e.g., *Il coniglio il pinguino lo tocca*/the rabbit the penguin it—CL touches) would be the only possible option with agentive verbs, as were all of the verbs used in the experiment.

(8) *a*-Topics in presence of lexical pre-verbal subjects



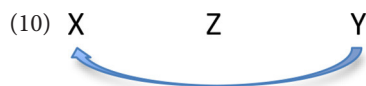
The realization of the preposed object topic as an *a*-Topic clearly correlates with the nature and position of the subject. This is shown by the fact that in some cases children used either a null subject or a post-verbal subject. Both options are grammatical in a null subject language like Italian, although the discourse conditions favored the overt and preverbal realization of the lexical subject, indeed the most widely adopted option by children. However, in those cases in which children opted for the null or post-verbal realization of the subject in the clause following the left dislocated object topic, the latter has been realized either as an *a*-Topic or as a simple Topic, with no *a* in an almost identical proportion. (9) Illustrates the distribution of *a*-Topics and simple Topics according to the nature and position of the subject:

(9) Comparing *a*-Topics and simple/O-Topics according to the nature (lexical or null) and position (pre-verbal or post-verbal) of the subject.



<sup>2</sup>Note that the lexical subject may either precede (as in example 2) or follow the *a*-Topic (e.g., compared to 2: *al pinguino il coniglio lo tocca*/to the penguin the rabbit him-CL touches); in the former case it also fills a left peripheral topic position, multiple topics being possible in Italian (Rizzi, 1997; Frascarelli and Hinterhölzl, 2007 a.o.); in the latter case it may either fill a Topic position lower than the *a*-Topic or be in the preverbal subject position within the clause (i.e., Spec/TP). The two possible orders are abbreviated as (S) preposed O (S) in the Figure in (8).

Why should the nature and position of the subject influence the realization of the Topic as an *a*-Topic? A principled reason can be assumed to be the origin of this influence. As discussed in detail in Belletti and Manetti (2017) CILD structures of the type in (2) under investigation display an object A'-dependency across an intervening lexical pre-verbal subject, in which both the preposed object and the subject are lexically restricted. According to the system developed in Friedmann et al. (2009), the notorious difficulty that children encounters with object A'-dependencies involving this intervention configuration—as in e.g. headed Object relative clauses with a pre-verbal lexical subject in the relative clause—may be accounted for in terms of the grammatical principle Relativized Minimality-RM expressed in featural terms, fRM (Rizzi, 1990, 2004; Starke, 2001; Grillo, 2008 for the proposal that the principle may also account for aspects of the agrammatic behavior in aphasia, on which see also Sheppard et al., 2015). According to the featural RM principle, in a configuration such as:



in which a dependency relation between the target position X and the origin position Y has to be established across an intervening Z<sup>3</sup>, such dependency cannot be established if X (target) and Z (intervener) share relevant features. In movement created dependencies, the relevant features are those triggering the displacement operation and attracting the relevant constituent in the target position. For instance, according to Friedmann et al. (2009), in headed object relatives the features attracting the relative head into the complementizer domain are a [+R] feature and a [+NP] feature. If a lexically restricted subject is present in the relative clause it also carries the [+NP] feature. Thus, by expressing the feature relations in set theoretic terms, the lexically restricted relative head (X), *il cane* of example (11), and the intervening lexically restricted subject (Z), *il gatto* in the same example (11), are in a relation of inclusion, with the feature [+NP] of the intervening lexical subject properly included within the feature set of the target.

- (11) *Il cane*    *che*    *il gatto*    *morde*    \_\_\_\_  
 the dog    that    the cat    bites  
 +R+NP       +NP  
 X                      Z                      Y

Indeed, if either the head of the object relative is not lexically restricted, as in the case Free object Relatives, or the subject of the relative clause is not lexically restricted as in the case in which it is a pronoun, object relatives are well understood by children, at the same level as subject relatives. This is the core experimental finding of Friedmann et al. (2009), which the

system captures through the proposal that there is development in the proper computation of the inclusion relation of the features which are relevant for the fRM principle. Further work has shown that the intersection relation of features relevant for the principle can be properly computed by young children (Belletti et al., 2012). Thus, for instance, illustrating once again with Italian, whereas a headed object relative like (11) is poorly comprehended by young children until a late age (still at 8-9 y.o. see Adani et al., 2010; Adani, 2011; Contemori and Belletti, 2014), object relatives in which the relative head and the subject of the relative clause mismatch in number are properly understood by children in a significantly higher proportion (Adani et al., 2010 for the relevant results on the mismatch configuration). This situation instantiates the intersection configuration, as (12) illustrates:

- (12) *Il cane*                      *che*    *i cavalli*    *rincorrono*    \_\_\_\_  
 the dog                      that    the horses    run after  
 +R+NP ±sing                      +NP + pl  
 X                                      Z                                      Y

Thus, according to the system in Belletti et al. (2012), grounded on Friedmann et al. (2009), given the four set theoretic relations, disjunction in relevant features is well-processed by both children and adults, identity is excluded by both (the core cases of classical RM, Rizzi, 1990); intersection is also well processed; in contrast, there is development in the proper computation of the inclusion relation of those features that the principle takes into account. The hypothesis is that such features are those that trigger syntactic movement, such as A'-movement into the left periphery of the clause. Thus, given this background, going back to the CILD structure under investigation here, the proposal can be made that resort to *a*-marking of the object Topic in the CILD structure containing a preverbal lexical subject is able to create an intersection relation between the feature composition of the target (X)—the left dislocated direct object—and of the intervener (Z)—the lexical subject. Under the assumption that *a*-Topics are associated with an affected interpretation of the topic a feature dubbed [+a] (affected, Belletti, 2017a) can be associated to an affected topic and a complementary feature [+u] to an unaffected one. The following intersection of relevant features illustrated in (13) is thus created, complying with the fRM principle (Belletti and Manetti, 2017, for all further relevant details):

- (13) *Il coniglio*                      *al pinguino*    *lo*    *tocca*  
 The rabbit                      to the penguin    him.Cl    touches  
 +Top +NP +u                      +Top+NP +a

In conclusion, it seems that a number of reasons may (have) contribute(d) to make *a*-Topics favored by young children in the results reviewed. Among them the following two play a crucial role:

- i. The fact that the preposed object, with which children establish an empathic relation is compatible with the (psychologically) affected interpretation associated with left peripheral *a*-Topics

<sup>3</sup>Where the intervention situation relevant to the principle is not linear but hierarchical, such that X c-commands Z, Z c-commands Y, and does not c-command X. The hierarchical nature of the relation is motivated on a vast amount of evidence in the theoretical linguistic literature. See Rizzi (2013) for recent discussion.

- ii. The fact that, in presence of a lexically expressed overt subject, resort to *a*-Topic effectively modulates the intervention problem posed by the syntactic configuration.

In contrast, frequency in the input of these structures appears to be an un-influential factor. As discussed, such structures do not really “exist” in standard Italian in the form widely adopted by the children. As the comparison in (9) strongly suggests, the crucial factor determining children's overextension in the use of *a*-Topics is the internal grammatical pressure of coping with a complex structure such as the one manifesting the hard intervention configuration<sup>4</sup>

## THE CASE OF *si*-CAUSATIVE PASSIVE IN CHILDREN'S PASSIVES

As mentioned in the introduction, sometimes children's answer to the question on the object of Belletti and Manetti's (2017) Experiment 1 that is reviewed here was a passive sentence (11%). Passive is a further appropriate type of answer given the experimental conditions, which corresponds in fact to the adults' widely adopted option (68%). See Belletti and Manetti (2017) for proposals on the possible reasons accounting for the difference between children and adults in the selection of the preferred answer to the elicitation question, which ultimately indicates that passive is not yet productively mastered at the ages under investigation as children's preferred answer in the same conditions were the CLD structures discussed in section The Case of *a*-topics in Children's CLDs<sup>5</sup>. Here, I would rather like to focus on the comparison between children and adults as for the types of passive utilized by the two groups in light of the issue of children's unexpected linguistic behavior under discussion here.

As is clearly illustrated in (14), children's and adults types of passives differ considerably: children exclusively selected the *si*-causative passive (all their 11% of passive answers were *si*-causative passives), whereas the most frequently utilized passives by adults were the periphrastic ones, copular or *venire* passive (49% out of their 68% of passive answers)<sup>6</sup>.

<sup>4</sup>Next to (13) the following order is also possibly realized in which the object *a*-Topic precedes the lexical subject (see footnote 2). The latter can thus fill either a peripheral topic position below the preposed object or the clause internal subject position. Nothing changes as for the intersection relation of relevant features:

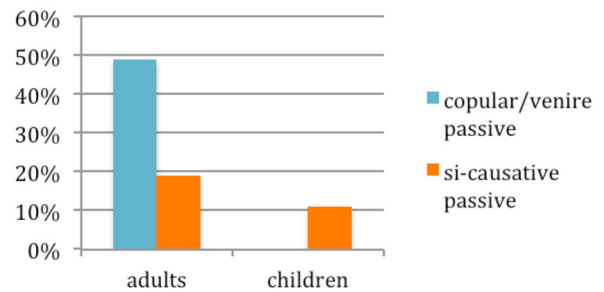
*Al re*                      *il bambino*                      *— lo sta pettinando —*  
+Top +NP +a (+Top) +NP +u  
(to the king            the boy                      him-CL is combing

As shown in Costa et al. (2014) for PP relatives in European Portuguese and Hebrew, the categorial distinction DP vs. PP cannot be considered the relevant distinctive factor.

<sup>5</sup>Adults did not produce CLDs in their answers (1%) as the passive answer took clear priority for them. This recalls the same adults' behavior found in previous work (Contemori and Belletti, 2014) whereby the production of relatives in the passive (i.e., Passive Object Relatives/PORs) in place of active object relatives was the overwhelmingly adopted answer by Italian speaking adults.

<sup>6</sup>More the latter than the former, as is natural with the actional verbs of the experiments all in a simple tense (present), the two fundamental conditions regulating *venire* passive.

## (14) Different types of passives produced by adults and children



The somewhat privileged status of *si*-causative passive in Italian speaking young children had also been found in previous experiments, using different techniques and eliciting different structures (e.g., *syntactic passive priming*, Manetti and Belletti, 2015; elicitation of object relatives through preference or picture description tasks, Contemori and Belletti, 2014). Let us now concentrate here on the significance of the sharp contrast revealed by figure (14) for the issue under investigation in this article.

We note first of all that the contrast in (14) cannot be due to children's sensitivity to the frequency of *si*-causative passive in their Italian input, since, as argued in Belletti (2017c), this type of passive is in fact rather rare compared to copular and *venire* passives in Italian<sup>7</sup>. Moreover, the computation involved in *si*-causative passive looks intuitively rather complex in that, beside including aspects of the computation also at play in copular and *venire* passive such as the *smuggling* operation moving a chunk of the verb phrase (Collins, 2005), it also involves one extra verb, the causative verb *fare* and the reflexive clitic *si* with the binding relation that its presence induces. However, far from being factors increasing the complexity of the computation, these grammatical properties are probably among those that make the *si*-causative passive more readily accessible to young children: on the one hand, the *smuggling* operation overtly triggered by the causative verb *fare* allows for a derivation in which intervention is properly eliminated (Manetti and Belletti, 2015) and on the other presence of the reflexive may constitute a further facilitating factor (Belletti, 2017b on the possible role of the reflexive, inducing a reflexive passive as a route to other types of passives; Belletti and Manetti, 2017 for further elaboration of these points). Thus, the robust access to *si*-causative passive that children have shown in this experiment, and which confirms previous independent results, indicates once again that children do not always do what they hear most. Furthermore, they also show early mastery of computations which are neither shorter nor simpler in any pre-theoretical sense, but which must count as such for their internal grammar.

<sup>7</sup>And indeed passive adults' answers to the elicitation question confirm that *si*-causative passive is not the passive that adults resort to most, as indicated in (14). Possibly their *si*-causative passive answers involved a causative/agentive type interpretation as it is generally the case for *si*-causative passives in adult standard Italian. It is hard to say whether the same interpretation is also necessarily at work in children's answers or whether their *si*-causative passive is just a form of passive with no necessary causative interpretation involved (as it is the mainly the case in e.g., standard French). See also the references quoted in the text on these points.

## CONCLUSIONS

The following conclusions can be drawn from the acquisition results reviewed here.

Grammatical and discourse related factors may sometimes lead children to systematically choose (the production of) structures which are only marginally present in the adult language. Overall, there does not seem to be any penalty for young children to access apparently complex and long(er) expressions *per se*, which can in fact sometimes be favored, as in the two cases reviewed. Both the *a*-Topics and the *si*-causative passives that children produced do involve longer expressions: a simple (object) topics without preposition *a*, is shorter than an *a*-Topic. Similarly, *copular* or *venire* passives, which do not involve the extra causative verb *fare* nor the reflexive clitic *si* with the implied binding relation are shorter and look simpler than *si*-causative passive. In both cases, however, the extra lexical elements may allow children to implement computations, which are in fact more readily accessible to their developing grammatical system.

Children thus end up displaying a grammatical behavior, which differs sharply from that of adults, as it happened in both cases considered here. Children's capacity to overextend given

syntactic structures thereby resulting in a grammatical creative behavior is the sign of an internal grammatical pressure, which does not necessarily require a rich input to be put into work<sup>8</sup>. The experimental conditions have succeeded in highlighting the children's grammatical creativity in newly identified contexts.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

The research presented here was funded in part by the European Research Council/ERC Advanced Grant 340297 SynCart—"Syntactic cartography and locality in adult grammars and language acquisition" which is here acknowledged. I also wish to thank Claudia Manetti, my coauthor of the work whose experimental results have inspired the reflections presented here.

<sup>8</sup>Nor does it appear to require special external conditions, such as e.g., multilingualism, contacts etc. The reported results all came from monolingual children speaking the same central Tuscan variety of Italian.

## REFERENCES

- Adani, F. (2011). Rethinking the acquisition of relative clauses in Italian: towards a grammatically based account. *J. Child Lang.* 38, 141–165. doi: 10.1017/S0305000909990250
- Adani, F., van der Lely, H. K., Forgiarini, M., and Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: a comprehension study with Italian children. *Lingua* 120, 2148–2166. doi: 10.1016/j.lingua.2010.03.018
- Belletti, A. (2017a). "On a-marking of object topics in the Italian left periphery," in *Beyond the Veil of Maya. From Sounds to Structures*, eds R. Petrosino, P. C. Cerrone and H. van der Hulst (Berlin: Mouton de Gruyter). (forthcoming).
- Belletti, A. (2017b). "Si as a route to passive in Italian," in *Linguistic Variation: Structure and Interpretation*, eds L. Franco and P. Lorusso (Berlin: Mouton de Gruyter). (forthcoming).
- Belletti, A. (2017c). "On the acquisition of complex derivations with related considerations on poverty of the stimulus and frequency," in *Syntactic Complexity from Acquisition*, ed E. Di Domenico (Newcastle: Cambridge Scholars Publishing), 28–48.
- Belletti, A., Friedmann, N., Brunato, D., and Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122, 1053–1069. doi: 10.1016/j.lingua.2012.02.007
- Belletti, A., and Manetti, C. (2017). *Topics and Passives in Italian-Speaking Children and Adults*. MS, University of Geneva-University of Siena.
- Belletti, A., and Rizzi, L. (1988). Psych verbs and Th-theory. *Nat. Lang. Linguist. Theory* 6, 291–352. doi: 10.1007/BF00133902
- Benincà, P. (1986). Il lato sinistro della frase italiana. *ATI J.* 47, 57–85.
- Benincà, P., and Poletto, C. (2004). "Topic, focus and V2. Defining the cp sublayers," in *The Structure of CP and IP*, ed L. Rizzi (New York, NY: Oxford University Press), 52–75.
- Berretta, M. (1989). Sulla presenza dell'accusativo preposizionale nell'italiano settentrionale: note tipologiche. *Vox Romanica* 48, 13–37.
- Bocci, G. (2013). *The Syntax-Prosody Interface. A Cartographic Perspective With Evidence from Italian*. Amsterdam; Philadelphia, PA: John Benjamins publishing Company.
- Collins, C. (2005). A smuggling approach to the passive in english. *Syntax* 8, 81–120. doi: 10.1111/j.1467-9612.2005.00076.x
- Contemori, A., and Belletti, A. (2014). Relatives and passive object relatives in Italian-speaking children and adults: intervention in production and comprehension. *Appl. Psycholinguist.* 35, 1021–1053. doi: 10.1017/S0142716412000689
- Costa, J., Friedmann, N., Silva, C., and Yachini, M. (2014). The boy that the chef cooked: Acquisition of PP relatives in European Portuguese and Hebrew. *Lingua* 150, 386–409. doi: 10.1016/j.lingua.2014.08.005
- Crain, S. (1991). Language acquisition in the absence of experience. *Behav. Brain Sci.* 14, 597–650. doi: 10.1017/S0140525X00071491
- Escandell-Vidal, V. (2009). Differential object marking and topicality the case of balearic catalan. *Stud. Lang.* 33, 832–885. doi: 10.1075/sl.33.4.02esc
- Frascarelli, M., and Hinterhölzl, R. (2007). "Types of topics in German and Italian," in *On Information Structure, Meaning and Form*, eds S. Winkler and K. Schwabe (Amsterdam; Philadelphia, PA: John Benjamins Publishing Company), 87–116.
- Friedmann, N., Belletti, A., and Rizzi, L. (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88. doi: 10.1016/j.lingua.2008.09.002
- Grillo, N. (2008). *Generalized Minimality: Syntactic Underspecification in Brocas Aphasia*. Utrecht: LOT Publications.
- Leonetti, M. (2004). Specificity and differential object marking in spanish. *Cat. J. Linguist.* 3, 75–114.
- Manetti, C. (2013). "On the production of passives in Italian: evidence from an elicited production task and a syntactic priming study with preschool children," in *Online Proceedings Supplement of the 37th Annual Boston University Conference on Language Development [BUCLD37]*, eds S. Baiz, N. Goldman, and R. Hawkes.
- Manetti, C., and Belletti, A. (2015). "Causatives and the acquisition of the Italian passive," in *Language Acquisition and Development Proceedings of GALA 2013*, eds C. Hamann and E. Ruigendijk (Newcastle: Cambridge Scholars Publishing), 282–298.
- Renzi, L. (1988). *Grande Grammatica Italiana di Consultazione, Vol. 1, La frase*. Bologna: Il Mulino.
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Rizzi, L. (1997). "The fine structure of the left periphery," in *Elements of Grammar*, eds L. Haegeman (Dordrecht: Kluwer), 281–337.
- Rizzi, L. (2004). "Locality and left periphery," in *Structures and Beyond The Cartography of Syntactic Structures, Vol 3*, ed A. Belletti (New York, NY: Oxford University Press), 223–251.

- Rizzi, L. (2013). Locality. *Lingua* 130, 169–186. doi: 10.1016/j.lingua.2012.12.002
- Rizzi, L., and Bocci, G. (2017). “The left periphery of the clause - primarily illustrated for Italian, 2nd Edn.,” in *The Wiley Blackwell Companion to Syntax, Second Edition*, eds M. Everaert and H. van Riemsdijk (Oxford: Blackwell Publishing).
- Sheppard, S. M., Walenski, M., Love, T., and Shapiro, L. P. (2015). The auditory comprehension of WH-questions in aphasia: support for the intervener hypothesis. *J. Speech Lang. Hear. Res.* 58, 781–797. doi: 10.1044/2015\_JSLHR-L-14-0099
- Snyder, W., and Hyams, N. (2015). “Minimality effects in children's passives,” in *Structures, Strategies and Beyond. Linguistik Aktuell/Linguistics Today*, eds E. Di Domenico, C. Hamann, and S. Matteini (Amsterdam; Philadelphia, PA: John Benjamins), 343–368.
- Starke, M. (2001). *Move Dissolves into Merge*. Doctoral dissertation, University of Geneva.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

**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 © 2017 Belletti. 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.



# The Peaceful Co-existence of Input Frequency and Structural Intervention Effects on the Comprehension of Complex Sentences in German-Speaking Children

Flavia Adani\*, Maja Stegenwallner-Schütz and Talea Niesel

Department of Linguistics, University of Potsdam, Potsdam, Germany

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Ernesto Guerra,  
University of Chile, Chile  
Chiara Gambi,  
University of Edinburgh,  
United Kingdom

### \*Correspondence:

Flavia Adani  
adani@uni-potsdam.de

### Specialty section:

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

**Received:** 30 April 2017

**Accepted:** 30 August 2017

**Published:** 29 September 2017

### Citation:

Adani F, Stegenwallner-Schütz M and  
Niesel T (2017) The Peaceful  
Co-existence of Input Frequency and  
Structural Intervention Effects on the  
Comprehension of Complex  
Sentences in German-Speaking  
Children. *Front. Psychol.* 8:1590.  
doi: 10.3389/fpsyg.2017.01590

The predictions of two contrasting approaches to the acquisition of transitive relative clauses were tested within the same groups of German-speaking participants aged from 3 to 5 years old. The input frequency approach predicts that object relative clauses with inanimate heads (e.g., *the pullover that the man is scratching*) are comprehended earlier and more accurately than those with an animate head (e.g., *the man that the boy is scratching*). In contrast, the structural intervention approach predicts that object relative clauses with two full NP arguments mismatching in number (e.g., *the man that the boys are scratching*) are comprehended earlier and more accurately than those with number-matching NPs (e.g., *the man that the boy is scratching*). These approaches were tested in two steps. First, we ran a corpus analysis to ensure that object relative clauses with number-mismatching NPs are not more frequent than object relative clauses with number-matching NPs in child directed speech. Next, the comprehension of these structures was tested experimentally in 3-, 4-, and 5-year-olds respectively by means of a color naming task. By comparing the predictions of the two approaches within the same participant groups, we were able to uncover that the effects predicted by the input frequency and by the structural intervention approaches co-exist and that they both influence the performance of children on transitive relative clauses, but in a manner that is modulated by age. These results reveal a sensitivity to animacy mismatch already being demonstrated by 3-year-olds and show that animacy is initially deployed more reliably than number to interpret relative clauses correctly. In all age groups, the animacy mismatch appears to explain the performance of children, thus, showing that the comprehension of frequent object relative clauses is enhanced compared to the other conditions. Starting with 4-year-olds but especially in 5-year-olds, the number mismatch supported comprehension—a facilitation that is unlikely to be driven by input frequency. Once children fine-tune their sensitivity

to verb agreement information around the age of four, they are also able to deploy number marking to overcome the intervention effects. This study highlights the importance of testing experimentally contrasting theoretical approaches in order to characterize the multifaceted, developmental nature of language acquisition.

**Keywords:** relative clauses, sentence comprehension, input frequency, number, animacy, language acquisition, German

## INTRODUCTION

Child language acquisition is a multifaceted process, which is likely to be influenced by several factors including structural rule learning, statistical learning, and social learning (e.g., Gervain and Mehler, 2010). The performance of children in experimental studies on complex sentences has often been used as a prism to infer which factors can be deployed to achieve an adult-like interpretation (e.g., Roeper, 2007). Among complex sentences, relative clauses have been used to test different language acquisition theories (Ambridge and Lieven, 2011).

The study presented in this paper was specifically designed to test, within the same participant groups, the predictions of what we will be calling the *input frequency* approach and the *structural intervention* approach. While the input frequency approach mainly concentrates on the distributional factors that influence children's early linguistic knowledge and its usage, the structural intervention approach focuses on grammatical mechanisms that may hinder or enhance the emergence of this knowledge. So far, the predictions of these approaches have been only tested in separate studies, using different participant groups and methods. Thus, the potential co-occurrence of frequency- and structure-driven effects, which might, simultaneously or successively, guide the performance of children during development can only be inferred indirectly.

In our study, we address the question of which factors support the comprehension of relative clauses by children from a new angle. First, a corpus study was conducted to identify whether the predictions of the structural intervention account, with respect to number dissimilarity effects, differ from the predictions of the input frequency account, regarding the frequency of number dissimilarities in the input. Next, we used a novel experimental design to draw a direct comparison of the predictions of the input frequency and structural intervention approaches, within the same participants and across different age groups, namely 3-, 4-, and 5-year-olds. To anticipate our findings, we were able to uncover that the effects predicted by input frequency and by the structural intervention approaches co-exist and that they both influence children's performance on transitive relative clauses, but in a way that is modulated by age. These effects hold at the group level but they are also reflected at the level of participant's individual performance.

The paper is organized as follows: First, the rationales behind the input frequency and the structural intervention approaches will be introduced and the existing studies on animacy and number dissimilarity (the two factors that are manipulated in our experiment) will be reviewed. Next, the hypotheses made by the two theoretical approaches will be tested by means of a corpus

study and an experimental study. A discussion of these results and of the co-existence of frequency and structural factors on the development of complex sentences will conclude the paper.

## The Input Frequency Approach

At the core of the input frequency approach lies the question of which environmental factors influence the emergence of children's early linguistic knowledge and its usage (Tomasello, 2000; Lieven, 2010). A few published studies have addressed this question with regard to the acquisition of relative clauses.

Based on the analysis of spontaneous speech data, Diessel and Tomasello (2000) proposed that, up to 3 years of age, English-speaking children's mastery of relative clauses is limited to structures that occur frequently in their own repertoires and which have a simple communicative function e.g., presentational constructions such as *Here's a tiger that's gonna scare him*. These sentences were analyzed as expressing a single proposition i.e., the tiger is going to scare him (cf. also Brandt et al., 2008 for converging evidence in the acquisition of German). At the same time, children's production of unequivocally fully-fledged relative clauses embedded in a main clause was mostly limited to subject-extracted relatives with intransitive verbs [e.g., *Is this something that turn(s) around?*]. Diessel and Tomasello (2000) showed that children begin to produce fully transitive subject and object relative clauses mostly between 4 and 5 years of age. The ability to correctly repeat mostly relative clauses with intransitive verbs was also found in English- and German-speaking 4-year-olds (Diessel and Tomasello, 2005), while the repetition of transitive relative clauses was significantly less accurate. Next, Kidd et al. (2007) designed a sentence repetition task where the properties of the sentences to be repeated reflected the distributional frequencies of these constructions in the input. The analysis of adult speech corpora showed that subject-extracted relative clauses tend to be more frequent than object-extracted relative clauses (Roland et al., 2007). Moreover, object relatives typically occur with an inanimate head noun and/or with a pronoun as embedded subject (e.g., *the car that she borrowed had a low tyre*) rather than with two animate NPs as verb arguments (e.g., *the cat that the dog is chasing is running very fast*). This pattern appeared to be fairly robust across (also typologically different) languages, such as English (Fox and Thompson, 1990), German (Mak et al., 2002), and Hebrew (Arnon, 2010)<sup>1</sup>. A number of adult sentence processing studies pointed toward a facilitation in object relative

<sup>1</sup>However, not for all these languages, the animacy and the pronoun constraints were evaluated simultaneously. Mak et al. (2002) only investigated animacy while Arnon (2010) limited her investigation to the distribution of embedded pronouns in child- and child-directed speech. Kidd et al. (2007) assessed both distributional constraints in child data.

clauses with inanimate head nouns (Traxler et al., 2002; Gennari and MacDonald, 2009; Wells et al., 2009) and with embedded pronominal subjects (Real and Christiansen, 2007), compared to object relative clauses with two animate full NPs. Based on these findings, Kidd et al. (2007) put forward the hypothesis that, if children's language processing system obeys the same constraints as the adult system, children should be able to repeat more faithfully object relative clauses with inanimate head nouns and embedded pronominal subjects. Indeed, this is what was found in English- and German-speaking 3- and 4-year-olds' productions. In English, the proportion of correctly repeated object relative clauses with an inanimate head and an embedded pronominal subject was similar to the proportion of correctly repeated subject relative clauses (~60% correct). In German, the proportion of correctly repeated, most frequently attested object relative clause structure was even higher than correctly repeated subject relative clauses. This discrepancy, however, seems to be rather due to the subject relatives, whose accuracy was surprisingly low (~20%) compared to object relatives (~60% correct). Similar accuracy patterns were obtained for comprehension by Brandt et al. (2009) by testing English- and German-speaking 3-year-olds using a referent selection task. In both languages, object relative clauses with two animate NPs were less accurate than object relative clauses with inanimate heads, which were, in turn, at least as accurate as subject relative clauses<sup>2</sup>. Independently of the work conducted within the input frequency approach, animacy effects during relative clause comprehension were also investigated in French-speaking 5- to 11-year-olds (Bentea et al., 2016) and Italian-speaking 9-year-olds (Arosio et al., 2011), showing converging results to the ones outlined above with younger children.

## The Structural Intervention Approach

Differently from the input frequency approach, the structural intervention approach aims at identifying which grammatical mechanisms may hinder or enhance the emergence of children's early linguistic knowledge (Guasti, 2002; Hyams and Orfitelli, 2017). With respect to the acquisition of relative clauses, Friedmann et al. (2009) argued that it is the structural similarity between the embedded subject and the head noun that hinders the comprehension of object relative clauses with two animate NPs, such as *Show me the cat that the dog is chasing*. The two NPs are structurally similar in the sense that they both contain an overt noun (*cat*, *dog*), a "lexical restriction" as Friedmann et al. called it. This overt noun on the embedded subject NP (*dog*) intervenes between the head noun (*cat*) and the position where this noun is interpreted as an object of the verb *chase*. This position and the one where the noun is pronounced as relative clause head are argued to be connected via syntactic movement. According to Friedmann et al. (2009), but cf. also Grillo (2009), the structurally similar, intervening subject NP disrupts the establishment of the movement dependency and the

correct interpretation of the sentence. Friedmann et al. (2009) tested Hebrew-speaking 3- to 5-year-olds (mean age 4;6) by comparing the production and comprehension of object relative clauses with two animate full NPs with those where only one argument is a full NP, such as object free relatives e.g., *Show me the who that the dog is chasing*, a well-formed sentence of Hebrew (as well as other conditions). Children were more accurate on those conditions where only one of the two verb arguments contained an overt noun (e.g., object free relatives), compared to object relatives with two full NPs. Hence, the prediction that the presence of an overt noun in the full NP hinders the comprehension of object relative clauses was borne out. By testing Italian- and English-speaking children, Adani et al. (2010, 2014) refined the notion of structural similarity taking into account grammatical features that are encoded within the NP, such as gender and number. As for English, Adani et al. (2014) showed that center-embedded subject as well as object relative clauses where the embedded NP and the head NP differ in terms of number (i.e., one is plural and the other is singular as in *The cat that is washing the goats/that the goats are washing has climbed onto the stool*) were understood significantly more accurately than the same structures without number dissimilarity (i.e., where both nouns are singular as in *The cat that is washing the goat/that the goat is washing has climbed onto the stool*). This result suggests that NP-internal features, like number, are relevant in the computation of structural intervention. Moreover, Adani et al. (2010) tested Italian center-embedded object relative clauses similar to the English study in three groups of Italian-speaking 5-, 7-, and 9-year-olds. The number dissimilarity effect was replicated, but Italian added the possibility to test also gender-marking. In contrast to number, gender dissimilarities yielded a significantly smaller facilitation effect. Hence, Adani et al. (2010) conclude that not all NP-internal features are equally relevant in the computation of structural intervention effects. The reduced facilitation of gender marking in the facilitation of object relative clauses in Italian was also tested in a subsequent study by Belletti et al. (2012), where Italian and Hebrew were compared. Subject- and object relative clauses of the type *Show me the dog that the goat is chasing* were tested separately in Hebrew and Italian in two groups of 3- to 5-year-olds (mean age 4;7 for each language). Italian and Hebrew crucially differ with respect to gender-marking: in Italian, it is only marked on nouns (e.g., *dog* is masculine and *goat* is feminine) while in Hebrew it is marked on the noun as well as on the verb via subject-verb agreement. Belletti et al. (2012) found that gender marking facilitates the comprehension of relative clauses in Hebrew but not (or, rather, to a much lesser extent) in Italian, similarly to what Adani et al. (2010) also found. Hence, Belletti et al. (2012) argued that a facilitation in structural intervention configurations only comes from features that are triggers of syntactic movement, typically inflected on the verb (e.g., number in English and in Italian, gender in Hebrew).

The hypothesis put forward by Belletti et al. (2012) is precisely the version of the structural intervention account that we will investigate in the present study and whose predictions will be compared to the predictions of the input frequency account (see Riches and Garraffa, 2017 for pursuing a similar goal but

<sup>2</sup>Both studies (Kidd et al. and Brandt et al.) tested, in addition to the animacy constraint, the pronoun constraint. This comparison yielded converging response patterns. But since our experiment focuses on the animacy constraint, the pronoun constraint will be not further discussed (but see also Arnon, 2010; Haendler et al., 2015).

focusing on different structures). The properties of German suit well these purposes as we know from previous studies that object relative clauses with inanimate heads are more frequent in the input (Mak et al., 2002), easier to imitate (Kidd et al., 2007) and to comprehend (Brandt et al., 2009) than object relative clauses with two animate NPs. Moreover, similarly to English and Italian, number agreement is overtly marked on verbs in German (Eisenberg, 2013). Coming to the predictions for our study, the input frequency approach predicts object relative clauses with an inanimate head and an animate embedded subject (OR:IN-AN, 1) to be more accurate than object relative clauses with two animate and singular NPs (OR:AN-AN, 2). The structural intervention approach predicts OR:AN-AN to be harder than object relative clauses with a singular head and a plural embedded NP (OR:SG-PL, 3)<sup>3</sup>.

1. The pullover that the man is scratching (OR:IN-AN)
2. The man that the boy is scratching (OR:AN-AN)
3. The man that the boys are scratching (OR:SG-PL)

In order to address further predictions of the two approaches, two types of subject relative clauses were tested as well. Based on the previous studies conducted in the spirit of the input frequency approach, OR:IN-AN are expected to be as accurate (or even more accurate) than subject relative clauses with two animate NPs (SR:AN-AN, 4). On the other hand, the structure intervention approach predicts the number marking facilitation to be specific for object relative clauses. Hence, when comparing SR:AN-AN with subject relative clauses with a singular head and a plural embedded NP (SR:SG-PL, 5), the structural intervention account predicts no difference between the two:

- 4 The man that is scratching the boy (SR:AN-AN)
- 5 The man that is scratching the boys (SR:SG-PL)

Crucially, however, it is not clear from the published literature whether a number facilitation in these contexts could also be predicted by the input frequency approach. According to the structural intervention account, object relative clauses with one singular NP and one plural NP are expected to be easier to interpret than object relative clauses with two verb arguments of the same number. Number is a movement-triggering feature and helps to reduce intervention between the moved head NP (the object) and the embedded NP (the subject). A facilitation in the same direction would be predicted by the input frequency account only if object relative clauses with one singular NP and one plural NP were more frequent in the input than object relative clauses with two verb arguments of the same number. To our knowledge, the question of how frequent relative clauses with number dissimilarity are (compared to relative clauses with number match), has not been examined yet in the existing literature. In order to set the basis for our experimental study, we report the data of a corpus study in which this question was addressed.

<sup>3</sup>Note: OR, object relative clause; SR, subject relative clause; AN-AN, two animate and singular NPs; IN-AN, one inanimate NP and one animate NP; SG-PL, one singular NP and one plural embedded NP.

## MATERIALS AND METHODS

### Corpus Study

In order to examine the input frequency of relative clauses with and without number mismatch between the head NP and the embedded NP, the speech of adults interacting with three German-speaking children from the CHILDES corpus (MacWhinney, 2000) was analyzed. The three corpora were those of Caroline (age range: 0;1–4;3), Kerstin (1;3–3;4), and Simone (1;9–4;0). All relative clauses containing the relative pronouns *der*, *die*, *das*, *den*, *wer*, *was*, *welcher*, *welches*, *welche*, and *wo* were extracted for a total of 307 utterances. All sentences were coded by the first author and subsequently checked by a native speaker with a linguistic background who was blind to the purpose of the analysis. All utterances were first classified as subject (SR,  $N = 134$ ) or object (OR,  $N = 173$ ) relative clauses. Among subject relative clauses, the ones containing the copular verb *sein* (to be) ( $N = 38$ , 28.5%), an intransitive verb ( $N = 59$ , 44%), or a reflexive verb ( $N = 5$ , 3.7%) were excluded from further analysis, leaving us with a total of 32 transitive subject relative clauses. For all transitive relative clauses, we analyzed whether the two NPs displayed (a) the same number; (b) a different number. Moreover, the two NPs were further analyzed in terms of their animacy properties, whether they are: (c) both animate; (d) the head noun is inanimate and the embedded NP is animate; (e) the head noun is animate and the embedded NP is inanimate; (f) both inanimate. These results are summarized in **Table 1**.

The distribution of NP types in object relative clauses was analyzed statistically with respect to two relevant comparisons: the occurrence of inanimate head nouns and the occurrence of number mismatch. Object relative clauses with inanimate heads were more frequent than object relative clauses with two animate NPs (binomial test,  $p < 0.001$ ) and both subject- and object relative clauses with number mismatch are rarer than their number match counterparts (binomial test for subject relatives,  $p = 0.007$ ; for object relatives,  $p < 0.001$ ).

To recap, some familiar and novel patterns emerge from this corpus study. First, over 70% of the subject relatives occurring in the child directed speech were either containing a copular verb or an intransitive verb. Although our analyses focused on transitive relative clauses only, the frequent occurrence of presumably simpler structures is noteworthy given previous claims that children's mastery of relative clauses before 3 years of age is limited to structures that occur frequently in the input, namely copular and intransitive relative clauses (Diessel and Tomasello, 2000, 2005). Second, both subject and object relative clauses with two animate NPs are rare in child directed speech and, within object relative clauses, significantly less frequent than object relative clauses with an inanimate head and an animate embedded subject, as also argued previously (Kidd et al., 2007; Brandt et al., 2009). We also found that object relative clauses with an animate head and an inanimate embedded subject are, overall, extremely rare and so are object relative clauses with two inanimate arguments. Most importantly, the novel information that emerges from this corpus study is that subject- and object relative clauses where the two

**TABLE 1** | Distribution of subject- and object- NP types in German child-directed speech.

| SR ( <i>N</i> = 32) |           |                     |           | OR ( <i>N</i> = 173) |           |                     |           |
|---------------------|-----------|---------------------|-----------|----------------------|-----------|---------------------|-----------|
| (a) Number match    |           | (b) Number mismatch |           | (a) Number match     |           | (b) Number mismatch |           |
| 24                  |           | 8                   |           | 115                  |           | 58                  |           |
| (c) AN-AN           | (d) IN-AN | (e) AN-IN           | (f) IN-AN | (c) AN-AN            | (d) IN-AN | (e) AN-IN           | (f) IN-IN |
| 6                   | 1         | 19                  | 6         | 21                   | 141       | 3                   | 7         |

SR, subject relative clauses; OR, object relative clauses; Number match, both NPs are singular or plural; Number Mismatch, one NP is singular and one is plural; AN-AN, both NP are animate; IN-AN, head NP is inanimate and embedded NP is animate; AN-IN, head NP is animate and embedded NP is inanimate; IN-IN, both NPs are inanimate.

NPs have different number are significantly rarer than the same structures where the two NPs display the same number. We can therefore conclude that, if a facilitation of object relative clauses with number dissimilarity will be attested in the experimental task, this is unlikely to be explicable on the basis of input frequency.

Coming to the experimental study, we put forward to following predictions, where “>” means “more accurate than” and “=” means “as accurate than as”:

- I Under the input frequency approach, OR:IN-AN > OR:AN-AN. Moreover, many studies testing this approach found that OR:IN-AN = SR:AN-AN;
- II Under the structural intervention approach, OR:SG-PL > OR:AN-AN. A derived but related prediction is that SR:AN-AN = SR:SG-PL;
- III The developmental trajectory will inform us whether the effects predicted by input frequency and structural intervention approaches co-exist or not;

## Experimental Task

### Participants

Seventy-three monolingual German-speaking children participated and were divided into three age groups: 23 three-year-olds (mean age 3;7, range 3;1–3;11), 25 four-year-olds (mean age 4;6, range 4;0–4;11), and 25 five-year-olds (mean age 5;4, range 5;0–5;11). Six additional children were tested but later excluded for one following reasons: lack of task completion (*N* = 1), difficulty in distinguishing the depicted characters (*N* = 1), color blindness, as indicated in the parental questionnaire (*N* = 1), failure to name three colors during the pre-test (*N* = 1), history of speech therapy (*N* = 2). This study was reviewed and approved by ethic committee at the University of Potsdam and it was carried out with parental written informed consent from all participants, in accordance with the Declaration of Helsinki. The study was piloted with a group of 4- to 5-year-olds and a group of adults (Adani, 2012) and later slightly modified in order to address the new research questions put forward in this paper. All children reported in this paper were tested with this updated version of the material.

### Material

The test sentences were transitive object- (OR, 6–8) and subject (SR, 9–10) relative clauses, for a total of 20 trials (four items per condition). Differently from English, both subject and object

relative clauses display the finite verb in clause final position in German, thus creating minimal pairs between the two extraction types, which are not confounded by overt word order differences. In addition to the test sentences, 16 relative clauses with intransitive verbs, e.g., (11), and prepositional phrases, e.g., (12), were included as fillers, for a total of 36 trials per list.

- 6 Welche Farbe hat der Mann, den die Jungen kratzen? (OR:SG-PL)  
Which color has the man whom the boys scratch
- 7 Welche Farbe hat der Mann, den der Junge kratzt? (OR:AN-AN)  
Which color has the man whom the boy scratches
- 8 Welche Farbe hat der Pulli, den der Mann kratzt? (OR:IN-AN)  
Which color has the pullover whom the man scratches
- 9 Welche Farbe hat der Mann, der den Jungen kratzt? (SR:AN-AN)  
Which color has the man who scratches the boy?
- 10 Welche Farbe hat der Mann, der die Jungen kratzt? (SR:SG-PL)  
Which color has the man who scratches the boys?
- 11 Welche Farbe hat der Mann, der trinkt? (Filler, intransitive)  
Which color has the man who drinks?
- 12 Welche Farbe hat der Pulli an der Wäscheleine? (Filler, prepositional)  
Which color has the pullover on the clothesline?

All trials were pseudo-randomized, with one filler after maximally three test sentences and it was never the case that two relative clauses of the same type followed each other. To neutralize order effects, the stimuli were administered in one order to half of the participants and in the reversed order to the other half. One stimulus list is reported in the Supplementary Material. **Figure 1** provides an example of three visual displays used for testing each of the five conditions.

Each visual display contained four pictures with characters of different colors. Only one picture displayed the target referent, e.g., a pullover in (8 and 12) or a man in the other conditions, performing the action expressed by the verb, e.g., scratching, and assuming the correct thematic role (either the agent or the patient). Each four-picture configuration was used to test one subject- and one object relative clause as well as two filler sentences, except in the animacy contrast, where subject relative clauses were not tested. Five additional characters in the scene were coded according to five non-target responses. In this paper, we decided not to report an analysis of non-target responses but

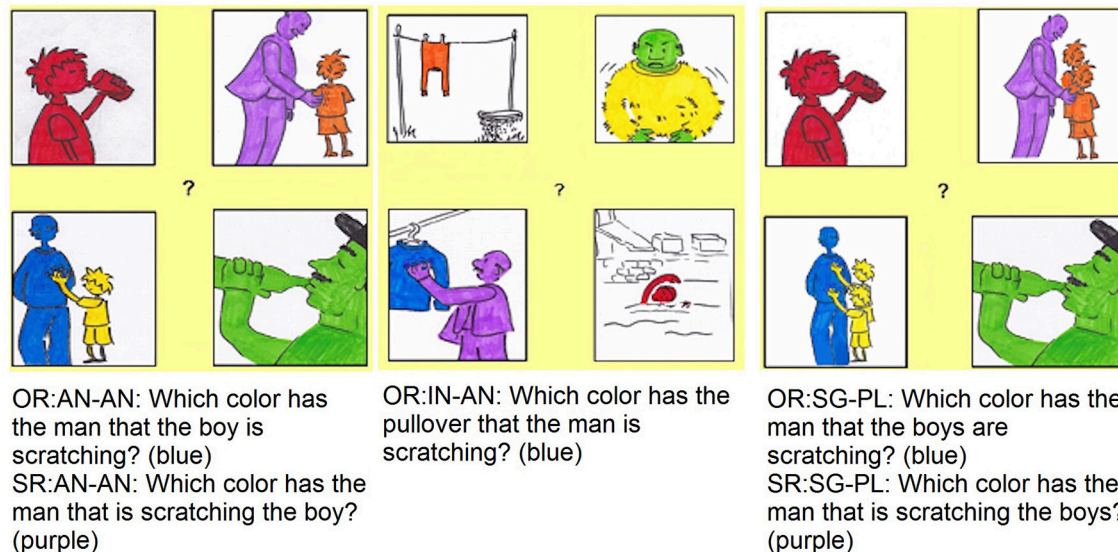


**FIGURE 1** | Example of the visual stimuli, related test sentences (English translation) and expected answer (in brackets). OR, object relative clause; SR, subject relative clause; AN-AN, two animate, singular NPs; IN-AN, one inanimate NP and one animate NP, both singular; SG-PL, one singular NP and one plural NP, both animate.

the interested reader can find an overview in the Supplementary Material.

## Procedure

The testing session started with a pre-test to ensure that children were familiar with the nouns: *Junge* (boy), *Baumstamm* (log), *Mann* (man), *Eimer* (bucket), *Gurt* (belt), *Schuh* (shoe), *Pulli* (pullover), *Frau* (lady), the verbs: *drücken* (to squeeze, to hug), *halten* (to carry, to hold), *kratzen* (to scratch), *tragen* (to carry, to hold) and the colors: *grün* (green), *lila* (purple), *blau* (blue), *orange* (orange), *gelb* (yellow), *rot* (red) used in the experiment. The lexical items, and in particular, the verbs, were chosen in such a way that either an animate or an inanimate noun could be a plausible subject. The interaction between the experimenter and the child was mediated by a snail puppet named Bala who introduced the sentence-picture-matching task in the form of a color naming game, inspired by Arnon (2010). The precise task instructions are reported in the Supplementary Material. At most one repetition of the trial was allowed. Before the experiment started, the participant was familiarized to the task by means of four practice trials. In addition to the color naming task, we administered a standardized non-word repetition task (Grimm et al., 2010), a selective attention test (Grob et al., 2009), a phonological memory test (Grob et al., 2009), and a sentence comprehension test (Sieg Müller et al., 2011). These tests were used as background measure for another study which is not reported in this paper. Each participant was tested individually by means of one or two sessions (depending on the child's individual pace and motivation) either in a quiet room of the day care or in a university laboratory. The whole testing lasted about 50 min for each participant, with breaks when needed. Children were generally engaged and happy to participate, and received a small toy as a reward.

## RESULTS

All responses were scored according to their correctness to the color naming question. Percentages of correct responses in filler sentences as well as test sentences were computed. Filler sentences were generally above 90% accurate, with some variation among 3-year-olds (92%), 4-year-olds (97%), and 5-year-olds (99%). Percentage of correct responses in each experimental condition is illustrated in **Figure 2**.

Visual inspection of **Figure 2** reveals that, as expected, subject relative clauses are overall more accurate than object relative clauses, when all conditions are taken together. The accuracy of object relative clauses varies considerably across conditions and age groups. The following statistical analysis was conducted to gain more insight as to which differences are statistically different across age groups.

We analyzed the proportion of correct responses with generalized linear mixed-models within the lme4 package (Bates et al., 2015b) in R (version 3.2.2; R Core Team, 2015). A logistic link function was used because of the binary dependent variable which could either have the value 1 (correct) or 0 (incorrect). We will present a statistical model that contains *group* and *condition* as categorical variable in its fixed effects structure. The three levels of group were contrasted in two steps: first, 4 year-olds (coded as 1) vs. 3 year-olds (coded as -1); next, 5 year-olds (coded as 1) vs. 4 year-olds (coded as -1). To address our research questions, we planned the following condition comparisons (the contrast coding is indicated in brackets):

- f1: OR:AN-AN (1) vs. OR:SG-PL (-1) → Is there a number dissimilarity facilitation?
- f2: OR:IN-AN (1) vs. OR:AN-AN (-1) → Is there an animacy dissimilarity facilitation?

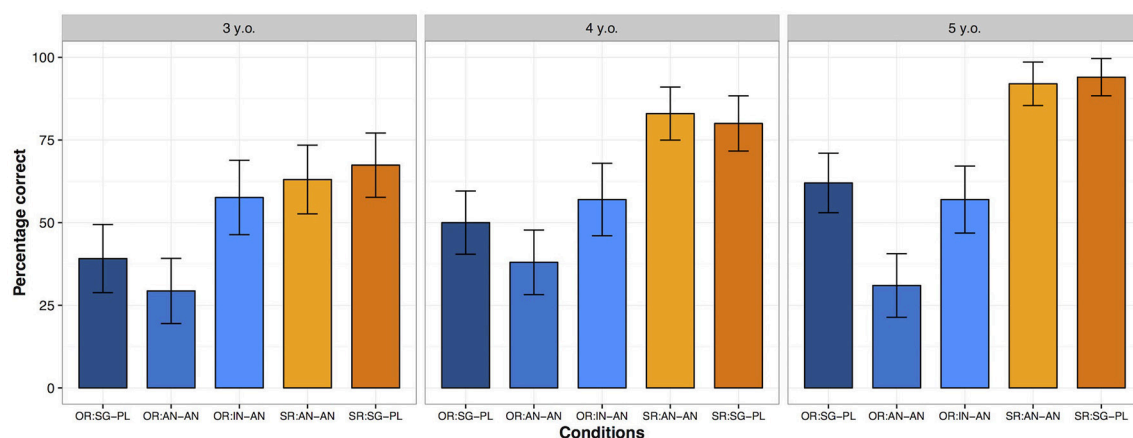


**FIGURE 2 |** Distribution of correct responses (in percentages) in each experimental condition, across age groups. The error bars represent 2 standard errors of the mean. OR, object relative clause; SR, subject relative clause; AN-AN, two animate and singular NPs; IN-AN, one inanimate NP and one animate NP, both singular; SG-PL, one singular NP and one plural NP, both animate.

f3: SR:AN-AN (1) vs. OR:IN-AN (−1) → Are OR with an inanimate head as easy as SR?

f4: SR:SG-PL (1) vs. SR:AN-AN (−1) → Is the number dissimilarity effect present also in SR?

The model tested for main effects of group and specific condition comparisons as well as interactions between group x condition. Following Bates et al. (2015a), we run a principle component analysis to identify the maximal model with only those random effects components that are supported by the data. The goodness-of-fit of nested alternatives of this model's random effects structure were evaluated with the *anova* function in R, based on the *p*-value associated to the chi-square-distributed likelihood ratio (Matuschek et al., 2017). After these checks, the random effects structure of the final model included varying subject intercepts and slopes for the comparison OR:IN-AN/OR:AN-AN and SR:AN-AN/OR:IN-AN as well as their correlation parameters. To sustain model convergence, we specified the *bobyqa* optimizer in the *glmer* function. The final model is the following:

```
m <- glmer(accuracy ~ group + f1 + f2 + f3 + f4 + group :
  f1 + group : f2 + group : f3 + group : f4 + (1 + f2
  + f3|subject_id), family = binomial, control
  = glmerControl(optimizer = "bobyqa"), data))
```

The output of this model is reported in **Table 2**. The statistically significant effects are highlighted in gray:

For interactions that were statistically significant in the main model, we nested the pairwise comparisons in order to explain the directions of the interactions. The significant pairwise comparisons are reported in the text below. The complete output of the two models with nested effects for the significant interactions are reported in the Supplementary Material.

We found significant main effects of group and of three out of the four pre-planned comparisons. The significant main

**TABLE 2 |** Model output.

| Fixed effects                        | Estimate | SE    | z-value | p-value |
|--------------------------------------|----------|-------|---------|---------|
| (Intercept)                          | 0.643    | 0.134 | 4.789   | <0.001  |
| 4 y.o. vs. 3 y.o.                    | 0.637    | 0.309 | 2.062   | 0.039   |
| 5 y.o. vs. 4 y.o.                    | 0.532    | 0.320 | 1.663   | 0.096   |
| OR:AN-AN vs. OR:SG-PL                | −1.007   | 0.208 | −4.851  | <0.001  |
| OR:IN-AN vs. OR:AN-AN                | 1.277    | 0.237 | 5.393   | <0.001  |
| SR:AN-AN vs. OR:IN-AN                | 1.547    | 0.257 | 6.013   | <0.001  |
| SR:SG-PL vs. SR:AN-AN                | 0.113    | 0.262 | 0.430   | 0.667   |
| 4 vs. 3 y.o. x OR:AN-AN vs. OR:SG-PL | −0.110   | 0.492 | −0.222  | 0.824   |
| 5 vs. 4 y.o. x OR:AN-AN vs. OR:SG-PL | −1.105   | 0.497 | −2.222  | 0.026   |
| 4 vs. 3 y.o. x OR:IN-AN vs. OR:AN-AN | −0.487   | 0.572 | −0.850  | 0.395   |
| 5 vs. 4 y.o. x OR:IN-AN vs. OR:AN-AN | 0.472    | 0.562 | 0.838   | 0.402   |
| 4 vs. 3 y.o. x SR:AN-AN vs. OR:IN-AN | 1.348    | 0.546 | 2.469   | 0.014   |
| 5 vs. 4 y.o. x SR:AN-AN vs. OR:IN-AN | 0.968    | 0.624 | 1.551   | 0.121   |
| 4 vs. 3 y.o. x SR:SG-PL vs. SR:AN-AN | −0.480   | 0.530 | −0.905  | 0.365   |
| 5 vs. 4 y.o. x SR:SG-PL vs. SR:AN-AN | 0.572    | 0.705 | 0.811   | 0.418   |

OR, object relative clause; SR, subject relative clause; AN-AN, two animate and singular NPs; IN-AN, one inanimate NP and one animate NP, both singular; SG-PL, one singular NP and one plural NP, both animate.

effect of group reveals that 3-year-olds perform significantly less accurately ( $M = 51\%$ ) than 4-year-olds ( $M = 62\%$ ), but the difference between 4- and 5-year-olds ( $M = 67\%$ ) does not reach significance.

The effect of the condition comparison OR:AN-AN vs. OR:SG-PL reveals that object relative clauses with number dissimilarity are more accurate ( $M = 51\%$ ) than object relative clauses with two animate, singular NPs ( $M = 33\%$ ). A significant interaction between group and OR:AN-AN vs. OR:SG-PL and subsequent pairwise comparisons reveal that, for 4- and especially for 5-year-olds, object relative clauses with number dissimilarity are significantly more accurate than object relative clauses with two animate, singular NPs (4 year-olds:  $\beta = -0.675$ ,

SE = 0.340,  $z = -1.987$ ,  $p = 0.047$ ; 5 year-olds:  $\beta = -1.780$ , SE = 0.369,  $z = -4.824$ ,  $p < 0.001$ ). In 3-year-olds, this effect shows a similar direction but it is not significant ( $\beta = -0.566$ , SE = 0.358,  $z = -1.581$ ,  $p = 0.114$ ). Moreover, the effect of condition comparison OR:IN-AN vs. OR:AN-AN and the absence of an interaction with group reveals that object relative clauses with an inanimate head and an animate embedded subject ( $M = 57\%$ ) are more accurate than object relative clauses with two animate NPs ( $M = 33\%$ ) and that this facilitation holds across all age groups. The effect of condition comparison SR:AN-AN vs. OR:IN-AN reveals that subject relative clauses with two animate, singular NPs ( $M = 80\%$ ) are overall more accurate than object relative clauses with an inanimate head ( $M = 57\%$ ). A significant interaction between group and SR:AN-AN vs. OR:IN-AN and subsequent pairwise comparisons reveal that subject relative clauses with two animate, singular NPs are more accurate than object relative clauses with an inanimate head for 4- and 5-year-olds (4-year-olds:  $\beta = 1.674$ , SE = 0.411,  $z = 4.074$ ,  $p < 0.001$ ; 5-year-olds:  $\beta = 2.642$ , SE = 0.500,  $z = 5.281$ ,  $p < 0.001$ ) but not for 3-year-olds ( $\beta = 0.326$ , SE = 0.367,  $z = 0.888$ ,  $p = 0.375$ ). There was no significant difference between subject relative clauses with two animate, singular NPs ( $M = 80\%$ ) and subject relative clauses with number dissimilarity ( $M = 81\%$ ).

## Individual Performance

In order to evaluate whether these group results reflect a response behavior that also holds at the individual level, we checked how many children named the correct color on at least 3 out of 4 trials per condition, within each age group. This corresponds to a probability of  $p < 0.05$  of providing the correct answer, assuming that each child has a 16.6% chance of naming the correct color. The chance level was fixed at 16.6% considering that the participant was expected to respond with one (correct) color out of six potential alternatives ( $100/6 = 16.6$ ). The results of this pass/fail analysis are reported in Table 3.

The response patterns of individual participants corroborate the group results along several dimensions. Starting with subject relative clauses, we found that the majority of the 3-year-olds already performed very accurately on both types (SR:AN-AN and SR:SG-PL). This rate increases in 4-year-olds and virtually all 5-year-olds performed correctly on both types of subject relative clauses. Moving to object relative clauses, we observe that the number of children who performed accurately on object relative clauses with two animate, singular NPs (OR:AN-AN) and with an inanimate head (OR:IN-AN) does not increase as a function of age. Rather, while only a restricted subgroup (12–20%) of children succeeds on OR:AN-AN, a larger subgroup of children (40–48%) performed accurately on the condition OR:IN-AN. What is nevertheless worth to emphasize is that these rates remain fairly steady across age groups. Finally, object relative clauses with number mismatch are correctly understood only by a restricted subgroup of 3-year-olds (22%). Crucially, for the research questions addressed in this study, the relative number of children who performed correctly on the OR with number dissimilarity was low in relation to the OR with an inanimate head (5 vs. 11, respectively). But by the age of 5 years, the relation between these two conditions flips its direction. While

**TABLE 3 |** Number (and percentages) of children who performed accurately on at least 3 out of 4 trials per condition.

| Age group           | OR:SG-PL (%) | OR:AN-AN (%) | OR:IN-AN (%) | SR:AN-AN (%) | SR:SG-PL (%) |
|---------------------|--------------|--------------|--------------|--------------|--------------|
| 3 y.o. ( $N = 23$ ) | 5 (22)       | 3 (13)       | 11 (48)      | 13 (57)      | 15 (65)      |
| 4 y.o. ( $N = 25$ ) | 9 (36)       | 5 (20)       | 10 (40)      | 22 (88)      | 20 (80)      |
| 5 y.o. ( $N = 25$ ) | 14 (56)      | 3 (12)       | 11 (44)      | 24 (96)      | 25 (100)     |

the number of children performing accurately on the OR with inanimate heads remains fairly steady, the number of passers on OR with number dissimilarity increases (11 vs. 14, respectively). The implications of the group and individual results will be discussed in the next section.

## DISCUSSION

A very thoroughly investigated question in the last decades' psycholinguistic research literature has been which type of information the human parser is relying on when processing filler-gap dependencies, of which relative clauses are a typical instance. Simplifying somehow a very multifaceted state-of-the-art, researchers have proposed several accounts, which capitalize on different sources of information that become crucial to achieve the correct interpretation of these sentences. For instance, some of these approaches emphasize the role of frequency (e.g., Gennari and MacDonald, 2009), some the role of memory resources (e.g., Lewis et al., 2006) and other the role of syntactic structure (e.g., De Vincenzi, 1990). Similar avenues have been pursued in the field of language acquisition. The main aim of this paper was precisely to bring together two of these approaches and test them systematically across 3-, 4-, and 5-year-olds.

The two approaches under discussion are the *input frequency* approach (Diessel and Tomasello, 2000, 2005; Kidd et al., 2007; Brandt et al., 2009; a.o) and the *structural intervention* approach (Friedmann et al., 2009; Grillo, 2009; Adani et al., 2010, 2014; Belletti et al., 2012; a.o). In order to evaluate the input frequency approach, we tested the prediction that object relative clauses with an inanimate head and an embedded animate NP are easier to interpret than object relative clauses with two animate NPs. Moreover, we also tested a derived prediction, which is often supported by the existing literature, that object relative clauses with an inanimate head are as easy as subject relative clauses with two animate NPs. On the other hand, in order to evaluate the structural intervention approach, we tested the prediction that object relative clauses with number dissimilarity are easier to interpret than object relative clauses with two singular NPs. Moreover, we have also tested the specificity of this prediction by comparing subject relative clauses with and without number dissimilarity.

In agreement with the input frequency approach, our corpus study converges with the claim that object relative clauses with an inanimate head and an embedded animate NP are the most frequent in child directed speech. Crucially, however, this

analysis also revealed that object relative clauses with number dissimilarity (one singular NP and one plural NP as verb arguments) are less frequent than object relative clauses with two animate, singular NPs. This suggests that a potential facilitation in the former condition could only be explained under the structural intervention approach and not in terms of input frequency.

Our experimental data reveal that object relative clauses with an inanimate head are more accurate than object relative clauses with two animate NPs, as the input frequency approach predicts. This response pattern is attested already in 3-year-olds and the facilitation sustains developmentally, as it is also attested in 4- and 5-year-olds. In 3-year-olds, however, the animacy dissimilarity is the main factor that enhances the accuracy on object relative clauses. These results are in contrast with the claim that only presentational, mono-clausal constructions and relative clauses with intransitive verbs are mastered by 3-year-olds, as Diessel and Tomasello (2000, 2005) proposed on the basis of a corpus study. Our data support the hypothesis that 3-year-olds are able to interpret transitive fully-fledged relative clauses, as long as they are of the frequently occurring type (Brandt et al., 2009).

Despite the early occurrence and the longitudinal robustness of an animacy effect, it is noted that the accuracy level in object relative clauses with an inanimate head and an embedded animate subject does not increase as the children grow older but remains around 57–58%. This response pattern is also reflected in the analysis of individual performances where we have noticed that around half of the group of 3-year-olds is already quite accurate in this condition, by performing on at least three out of four trials correctly. However, about half of the 5-year-olds still has some difficulties in interpreting these sentences fully correctly. At this point, we can only speculate that the animacy contrast information is immediately accessible even to very young children but that the successful deployment of this information to correct sentence interpretation is based on the application of some top-down, shallow processing heuristic rather than a bottom-up, deep processing of the sentence in the adult-like sense. Based on these results, we suggest that the sensitivity to input frequency information is not subject to development, meaning that it is available from very early on but it does not increase as the child's cognitive and linguistic development progresses.

Coming to the second prediction of the input frequency approach, namely that object relative clauses with an inanimate head can become as easy as subject relative clauses, our data do not support this prediction. We have found that 4- and 5-year-olds are still significantly more accurate on subject relative clauses than on the frequent object relatives with inanimate heads. This difference is not attested in 3-year-olds, though. However, a similar performance on these two conditions in 3-year-olds has more to do with the low accuracy of subject relative clauses in the youngest group. For this reason, the difference with object relative clauses with an inanimate head fades away. This finding is similar to the one reported by Kidd et al. (2007) where 3-year-olds were only able to repeat faithfully 20% of

the prompted structures. This apparent “difficulty” with subject relative clauses with two animate NPs could be linked to the lack of an animacy contrast and, as such, the impossibility for 3-year-olds to apply the above mentioned heuristic as successfully as they do with object relative clauses with inanimate heads. In the older groups, however, an advantage for subject relative clauses over object relative clauses with inanimate heads may signal that 4-year-olds start to provide a deeper, fully-fledged syntactic analysis of these structures and the interpretation of a subject relative clause does not pose a challenge. Another related observation that we leave open to future research concerns the co-occurrence of several distributional constraints that may incrementally support the comprehension of frequent object relative clauses. In this study, we have only manipulated the animacy constraint but, as noted in the introduction, the presence of an embedded pronominal subject may also play a crucial role in modulating the ease of object relative clauses. It could be that object relative clauses become as accurate as subject relative clauses only when both the animacy and the pronominal constraints are satisfied (cf. also the discussion in Arnon, 2010). We will now move to the discussion of effects related to structural intervention.

Coming to the predictions of the structural intervention approach, we found that number dissimilarity enhances the correct interpretation of object relative clauses, when all age groups are taken together. Moreover, this effect emerges in 4-year-olds but it becomes stronger in 5-year-olds. This is in line with the literature that has tested the structural intervention approach in children with a mean age of 4;6 or older ones. In all these studies, a dissimilarity of features that are triggers of syntactic movement enhanced object relative clause comprehension. These features are, for instance, number dissimilarity in Italian and English (Adani et al., 2010, 2014), gender dissimilarity in Hebrew (Belletti et al., 2012) and potentially in other languages in which subject-verb agreement is marked on the verb. There is independent evidence coming from research on subject-verb agreement using implicit measures (eye-movements) showing that between 3- and 5-years of age German-speaking children are fine-tuning their sensitivity to agreement information as well as to its violations. Brandt-Kobe and Höhle (2010) showed that 3-year-olds take advantage of the information on the verb inflection to identify the correct agreeing subject NP. However, this ability was only evident in the eye-gazes of the children but not in the explicit (pointing) responses. Considering the explicit nature of our task, in which children were asked to name the color of the relative clause head noun referent, and the fact that our test sentences are more complex than the declarative sentences tested by Brandt-Kobe and Höhle, it is not surprising that the number dissimilarity facilitation only emerges in the group of 4-year-olds. The apparent similarity between Brandt-Kobe and Höhle's data and the data of the present study suggests that the number effect might emerge already in 3-year-olds, when they are tested implicitly, for instance, measuring their eye-gazes. Nevertheless, what the explicit version of our task already shows is that, everything else being equal, there is an earlier advantage for the animacy dissimilarity over the number dissimilarity.

Coming to the second prediction of the structural intervention approach, we do find that subject relative clauses with number dissimilarity are not different from subject relative clauses without number dissimilarity. This result is in line with the predictions of the structural intervention approach, thus supporting the claim that these effects occur specifically in the intervention-triggering contexts. It may be worth to notice that the lack of statistical difference is not a consequence of an at-ceiling performance on subject relatives in general. Rather, each age group performed very similarly on the two subject relative conditions. As we have already discussed, their accuracy is quite low in 3-year-olds (SR:AN-AN: 63% correct; SR:SG-PL: 67%) and it increases gradually in 4-year-olds (SR:AN-AN: 83%; SR:SG-PL: 80%) and in the 5-year-olds (SR:AN-AN: 92%; SR:SG-PL: 94%).

The conclusions that can be drawn from our study are in agreement with most of the studies published by researchers that work with the input frequency approach and the structural intervention approach. The step forward that our study makes is to compare these two approaches directly, within the same children and across relatively large samples of participants belonging to different age groups. By doing so, we have discovered that input frequency and structural intervention effects co-exist and that the emergence of these effects is modulated by age. In all age groups, the animacy mismatch appears to explain children's performance, thus, showing that the comprehension of frequent object relative clauses is enhanced. These results reveal a sensitivity to animacy mismatch already in 3-year-olds and show that animacy is initially deployed more reliably than number to interpret relative clauses correctly. Once children fine-tune their sensitivity to verb agreement information around age four, they are also able to deploy number marking to overcome the intervention effect. Future avenues of investigation that our study opens up are the use of implicit measures, for instance eye-tracking, to shed more light on the abilities of 3-year-olds, who might not be able to cope with the explicit task demands as efficiently as the older children do. Our study also highlights the importance of comparing predictions of different developmental approaches in combination with cross-sectional

data to gain detailed insight into the dynamics of the acquisition process.

## AUTHOR CONTRIBUTIONS

FA and TN conceived the study and prepared the material; FA and MS acquired and analyzed the data. FA, MS and TN drafted, revised, and approved the final manuscript.

## FUNDING

This recruitment and testing of participants was partially supported by the DFG grant AD 408/1-1 to FA, which is gratefully acknowledged.

## ACKNOWLEDGMENTS

We want to express our gratitude to all children and their families whose participation made this study possible. We thank Jule Bergt for drawing the pictures. The data analyzed in the corpus study were part of Anna Kämpfner's BSc thesis at the University of Potsdam, which was jointly supervised by Yair Haendler and FA. We thank Anna Walther, Jasmin Biel, Sabine Peiffers, Leonie Lampe, Philippa Hildebrandt and Sophie Gruhn for their assistance with the data collection. We are also very grateful to Tom Fritzsche, Marinella Carminati, and Philippe Prévost for their comments on previous version of this work and to Leonie Lampe and Tom Fritzsche for their assistance with the corpus study. All remaining errors are, of course, our own. We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01590/full#supplementary-material>

## REFERENCES

- Adani, F. (2012). "Some notes on the acquisition of relative clauses: new data and open questions," in *ENJOY LINGUISTICS! Papers Offered to Luigi Rizzi on the Occasion of his 60th Birthday*, eds V. Bianchi and C. Chesì (Siena: CISCL Press).
- Adani, F., Forgiarini, M., Guasti, M. T., and van der Lely, H. K. J. (2014). Number dissimilarities facilitate the comprehension of relative clause in children affected by (Grammatical) specific language impairment. *J. Child Lang.* 41, 811–841. doi: 10.1017/S0305000913000184
- Adani, F., van der Lely, H. K. J., Forgiarini, M., and Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: a comprehension study with Italian children. *Lingua* 120, 2148–2166. doi: 10.1016/j.lingua.2010.03.018
- Ambridge, B., and Lieven, E. (2011). *Child Language Acquisition. Contrasting Theoretical Approaches*. New York, NY: Cambridge University Press.
- Arnon, I. (2010). Rethinking child difficulty: the effect of NP type on children's processing of relative clauses in Hebrew. *J. Child Lang.* 37, 27–57. doi: 10.1017/S030500090900943X
- Arosio, F., Guasti, M. T., and Stucchi, N. (2011). Disambiguating information and memory resources in children's processing of Italian relative clauses. *J. Psycholinguist. Res.* 40, 137–154. doi: 10.1007/s10936-010-9160-0
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv:1506.04967*
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Belletti, A., Friedmann, N., Brunato, D., and Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122, 1053–1069. doi: 10.1016/j.lingua.2012.02.007
- Bentea, A., Durrleman, S., and Rizzi, L. (2016). Refining intervention: the acquisition of featural relations in object A-bar dependencies. *Lingua* 169, 21–41. doi: 10.1016/j.lingua.2015.10.001

- Brandt, S., Diessel, H., and Tomasello, M. (2008). The acquisition of German relative clauses: a case study\*. *J. Child Lang.* 35, 325–348. doi: 10.1017/S0305000907008379
- Brandt, S., Kidd, E., Lieven, E., and Tomasello, M. (2009). The discourse bases of relativization: an investigation of young German and english-speaking children's comprehension of relative clauses. *Cogn. Linguist.* 20, 539–570. doi: 10.1515/COGL.2009.024
- Brandt-Kobe, O.-C., and Höhle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: the case of verb inflection in language acquisition. *Lingua* 120, 1910–1925. doi: 10.1016/j.lingua.2010.02.008
- De Vincenzi, M. (1990). *Syntactic Parsing Strategies in Italian: the Minimal Chain Principle*, Vol. 12, Dordrecht; Boston, MA: Kluwer Academic Publishers.
- Diessel, H., and Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cogn. Linguist.* 11, 131–151. doi: 10.1515/cogl.2001.006
- Diessel, H., and Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language* 81, 882–906. doi: 10.1353/lan.2005.0169
- Eisenberg, P. (2013). *Grundriss der Deutschen Grammatik*. Weimar: J.B. Metzler.
- Fox, B. A., and Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in english conversation. *Language* 66, 297–316. doi: 10.2307/414888
- Friedmann, N., Belletti, A., and Rizzi, L. (2009). Relativized relatives. Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88. doi: 10.1016/j.lingua.2008.09.002
- Gennari, S. P., and MacDonald, M. C. (2009). Linking production and comprehension processes: the case of relative clauses. *Cognition* 111, 1–23. doi: 10.1016/j.cognition.2008.12.006
- Gervain, J., and Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annu. Rev. Psychol.* 61, 191–218. doi: 10.1146/annurev.psych.093008.100408
- Grillo, N. (2009). Generalized minimality: feature impoverishment and comprehension deficits in agrammatism. *Lingua* 119, 1426–1443. doi: 10.1016/j.lingua.2008.04.003
- Grimm, H., Aktas, M., and Frevert, S. (2010). *SETK 3-5. Sprachentwicklungstest für Drei- bis Fünfjährige Kinder (2. Auflage). Diagnose von Sprachverarbeitungsfähigkeiten und Auditiven Gedächtnisleistungen*. Göttingen: Hogrefe.
- Grob, A., Meyer, C. S., and Hagmann-von Arx, P. (2009). *Intelligence and Development Scales (IDS). Intelligenz- und Entwicklungsskalen für Kinder von 5-10 Jahren*. Bern: Huber.
- Guasti, M. T. (2002). *Language Acquisition: the Growth of Grammar*. Cambridge, MA: MIT Press.
- Haendler, Y., Kliegl, R., and Adani, F. (2015). Discourse accessibility constraints in children's processing of object relative clauses. *Front. Psychol.* 6:860. doi: 10.3389/fpsyg.2015.00860
- Hyams, N., and Orfitelli, R. (2017). "The acquisition of syntax," in *Handbook for Psycholinguistics*, eds H. Cairns and E. Fernandez (Cambridge, MA: Wiley/Blackwell Publishers).
- Kidd, E., Brandt, S., Lieven, E., and Tomasello, M. (2007). Object relative made easy: a cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Lang. Cogn. Process.* 22, 860–897. doi: 10.1080/01690960601155284
- Lewis, R. L., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007
- Lieven, E. (2010). Input and first language acquisition: evaluating the role of frequency. *Lingua* 120, 2546–2556. doi: 10.1016/j.lingua.2010.06.005
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, 3rd Edn.* Mahwah, NJ: Lawrence Erlbaum Associates.
- Mak, W. M., Vonk, W., and Schriefers, H. (2002). The Influence of animacy on relative clause processing. *J. Mem. Lang.* 47:50. doi: 10.1006/jmla.2001.2837
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>
- Real, F., and Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *J. Mem. Lang.* 57, 1–23. doi: 10.1016/j.jml.2006.08.014
- Riches, N. G., and Garraffa, M. (2017). A discourse account of intervention phenomena: an investigation of interrogatives. *Glossa* 2:74. doi: 10.5334/gjgl.100
- Roeper, T. (2007). *The Prism of Grammar*. Cambridge: MIT Press.
- Roland, D., Dick, F., and Elman, J. L. (2007). Frequency of basic english grammatical structures: a corpus analysis. *J. Mem. Lang.* 57, 348–379. doi: 10.1016/j.jml.2007.03.002
- Siegmüller, J., Kauschke, C., van Minnen, S., and Bittner, D. (2011). *Test zum Satzverstehen von Kindern. Eine Profitorientierte Diagnostik der Syntax*. München: Elsevier.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* 74, 209–253. doi: 10.1016/S0010-0277(99)00069-4
- Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing of subject and object relative clauses: evidence from eye movements. *J. Memory Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836
- 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

**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 © 2017 Adani, Stegenwallner-Schütz and Niesel. 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.



# Comprehension of Subject and Object Relative Clauses in a Trilingual Acquisition Context

Angel Chan<sup>1,2,3\*</sup>, Si Chen<sup>1</sup>, Stephen Matthews<sup>4,5</sup> and Virginia Yip<sup>4</sup>

<sup>1</sup> Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong, China, <sup>2</sup> Speech Therapy Unit, Hong Kong Polytechnic University, Hong Kong, China, <sup>3</sup> Hong Kong Polytechnic University-Peking University Research Centre on Chinese Linguistics, Hong Kong, China, <sup>4</sup> Childhood Bilingualism Research Centre, The Chinese University of Hong Kong, Hong Kong, China, <sup>5</sup> Department of Linguistics, The University of Hong Kong, Hong Kong, China

Chinese relative clauses (RCs) have word order properties that are distinctly rare across languages of the world; such properties provide a good testing ground to tease apart predictions regarding the relative complexity of subject and object RCs in acquisition and processing. This study considers these special word order properties in a multilingual acquisition context, examining how Cantonese(L1)-English(L2)-Mandarin(L3) trilingual children process RCs in two Chinese languages differing in exposure conditions. Studying in an English immersion international school, these trilinguals are also under intensive exposure to English. Comparisons of the trilinguals with their monolingual counterparts are made with a focus on the directionality of cross-linguistic influence. The study considers how various factors such as language exposure, structural overlaps in the target languages, typological distance, and language dominance can account for the linguistic abilities and vulnerabilities exhibited by a group of children in a trilingual acquisition context. Twenty-one trilingual 5- to 6-year-olds completed tests of subject- and object- RC comprehension in all three languages. Twenty-four age-matched Cantonese monolinguals and 24 age-matched Mandarin monolinguals served as comparison groups. Despite limited exposure to Mandarin, the trilinguals performed comparable to the monolinguals. Their Cantonese performance uniquely predicts their Mandarin performance, suggesting positive transfer from L1 Cantonese to L3 Mandarin. In Cantonese, however, despite extensive exposure from birth, the trilinguals comprehended object RCs significantly worse than the monolinguals. Error analyses suggested an English-based head-initial analysis, implying negative transfer from L2 English to L1 Cantonese. Overall, we identified a specific case of bi-directional influence between the first and second/third languages. The trilinguals experience facilitation in processing Mandarin RCs, because parallels and overlaps in both form and function provide a transparent basis for positive transfer from L1 Cantonese to L3 Mandarin. On the other hand, they experience more difficulty in processing object RCs in Cantonese compared to their monolingual peers, because structural overlaps with competing structures from English plus intensive exposure to English lead to negative transfer from L2 English to L1 Cantonese. The findings provide further evidence that head noun assignment in object RCs is especially vulnerable in multilingual Cantonese children when they are under intensive exposure to English.

**Keywords:** child second and third language acquisition, cross-linguistic influence, input conditions, structural overlaps, typological distance, Cantonese, Mandarin, English

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Anouschka Foltz,  
Bangor University, United Kingdom  
Kathy Hirsh-Pasek,  
Temple University, United States

### \*Correspondence:

Angel Chan  
angel.ws.chan@polyu.edu.hk

### Specialty section:

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

**Received:** 20 April 2017

**Accepted:** 06 September 2017

**Published:** 06 October 2017

### Citation:

Chan A, Chen S, Matthews S and  
Yip V (2017) Comprehension of  
Subject and Object Relative Clauses  
in a Trilingual Acquisition Context.  
Front. Psychol. 8:1641.  
doi: 10.3389/fpsyg.2017.01641

## INTRODUCTION

Relative clauses have been intensely investigated in language typology, acquisition and processing for decades. Chinese relative clauses have word order properties that are otherwise rare across the languages of the world. Given these special word order properties, Chinese languages are important in debates regarding acquisition and processing of RCs because they allow researchers to tease apart predictions regarding the relative complexity of subject vs. object RCs. Moreover, relative clauses in Cantonese and Mandarin differ enough for there to be language-specific effects on acquisition (Chan et al., 2011).

In this study, we take on a new perspective by considering these special word order properties in a multilingual acquisition context, examining how Cantonese(L1)-English(L2)-Mandarin (L3) trilingual children process relative clauses in two Chinese languages acquired under different exposure conditions. The trilinguals come from Hong Kong middle class families where they are exposed to Cantonese as first language in the family and community from birth, and to Mandarin at school for only 200 min per week. Being educated in an English immersion international school, the trilinguals acquire these two Chinese languages under intensive exposure to English as a second language. Comparisons of the trilingual children with their monolingual counterparts are made with a focus on the directionality of cross-linguistic influence. The study considers language exposure, structural overlaps in the target languages, typological distance, perceived language distance, and language dominance as factors leading multilingual children to experience facilitation in one instance and competing analyses in another, when processing relative clauses (Chan et al., 2011; Kidd et al., 2015).

The study is novel in a number of ways. First, it is the first experimental study of relative clause comprehension in Cantonese-English-Mandarin trilingual children. Second, we demonstrate a specific instance of bi-directional influence between first and second/third languages in this syntactic domain in a trilingual acquisition context. In particular, we argue for forward positive transfer from L1 Cantonese to L3 Mandarin and reverse negative transfer from L2 English to L1 Cantonese taking place within a single grammatical domain in this group of trilinguals. L2-to-L1 transfer has been documented in a number of studies involving a variety of language pairs, although to date, a majority of the studies feature adult second language acquisition in a largely European language context (e.g., Cook, 2003; Dussias and Sagarra, 2007; Morett and MacWhinney, 2013; but see also Liu et al., 1992; Su, 2001 involving Mandarin and English in adult second language acquisition). Third, this study features the acquisition of Chinese under strong English influence, a phenomenon that is increasingly common among not only children in Hong Kong who are being educated in an international school curriculum, but also relevant to a significant number of Chinese immigrant or adopted children around the world who are typically exposed to a Chinese language at home and grow up in an English-speaking country where they acquire the English language of the speech community simultaneously

or successively. These multilingual children form a significant emerging group facing the challenge of preserving Chinese as their heritage language and acquiring English as the mainstream language of the community and/or school in which they grow up.

## Relative Clauses in Cantonese, Mandarin, and English

While English and Chinese share the basic word order SVO, they differ in that relative clauses (RCs) are consistently placed before the head noun in Chinese. See (1) and (2) for an example of a subject RC, and (3) and (4) for an example of an object RC in Cantonese and Mandarin respectively. In fact, pre-nominal RCs plus SVO main clause word order is a rare combination cross-linguistically (Dryer, 2013).

**Cantonese subject RC** (CL: classifier; SFP: sentence final particle):

- (1) [RC<sub>i</sub> 錫 公雞] [head noun 嗰 隻 老鼠]<sub>i</sub>  
 sek3 gung1gai1 go2 zek3 lou5syu2  
 kiss chicken that CL mouse  
 喺 邊度 呀?  
 hai2 bin1dou6 aa3  
 is where SFP  
 "Where's the mouse that kisses the chicken?"

**Mandarin subject RC:**

- (2) [RC<sub>i</sub> 親 公雞] 的 [head noun 老鼠]<sub>i</sub> 在 哪裡?  
 qing gongji de laoshu zai nali  
 kiss chicken de mouse is where  
 "Where's the mouse that kisses the chicken?"

**Cantonese object RC:**

- (3) [RC 老鼠 錫<sub>j</sub>] [head noun 嗰 隻 公雞]<sub>j</sub>  
 lou5syu2 sek3 go2 zek3 gung1gai1  
 mouse kiss that CL chicken  
 喺 邊度 呀?  
 hai2 bin1dou6 aa3  
 is where SFP  
 "Where's the chicken that the mouse kisses?"

**Mandarin object RC:**

- (4) [RC 老鼠 親<sub>j</sub>] 的 [head noun 公雞]<sub>j</sub> 在 哪裡?  
 laoshu qing de gongji zai nali  
 mouse kiss de chicken is where  
 "Where's the chicken that the mouse kisses?"

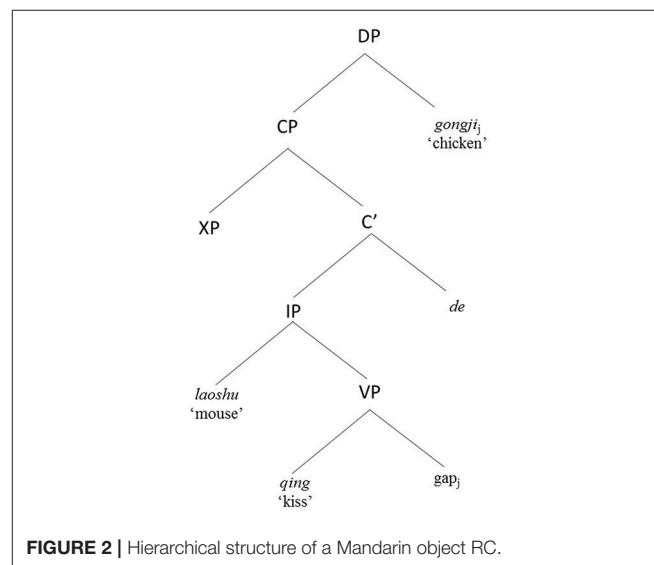
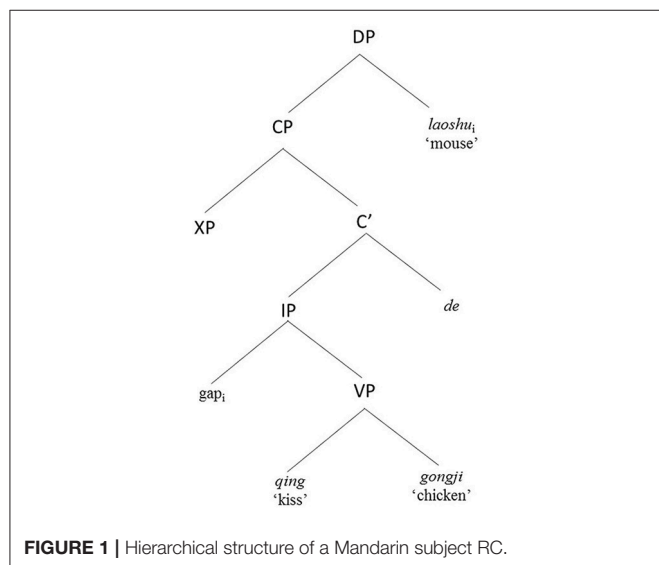
As illustrated by examples (1) to (4), placing the RC before the head noun results in Chinese subject RCs having non-canonical VOS word order and a longer linear distance between filler and gap, while Chinese object RCs match the canonical SVO word order and have a shorter linear filler-gap distance. These structural configurations result in competing processing demands described as follows. On the one hand, Chinese subject RCs are less costly to process due to general subject prominence

based on functional notions such as topicality: given that relative clauses describe the referent of their head noun and a clause's subject constitutes the default topic, it is less effortful to construe a RC as being about its default topic (the subject) than to construe it as being about some other item (Keenan and Comrie, 1977; Kim and O'Grady, 2015). From a formalist perspective, Chinese subject RCs are also easier to process in terms of lack of structural intervention in a hierarchical structure (Hu et al., 2015a,b). Along the lines of the locality principle of Relativized Minimality (Rizzi, 1990), a local relation between X (the relative head noun in the case of RCs) and Y (the copy of the moved relative head noun in the gap position) cannot hold if there is an intervener, Z, which is of the same structural type as X, and can be a potential candidate for the relation. In Chinese subject RCs like **Figure 1**, there is no structural intervener between the relative head (*laoshu* "mouse") and its copy in the gap position. However, in Chinese object RCs like **Figure 2**, the embedded subject (*gongji* "chicken") intervenes between the relative head (*laoshu* "mouse") and its copy in the gap position, and qualifies as a potential candidate for the local relation. This makes the correct computation of the local relation more complex to resolve for children when they process Chinese object RCs. On the other hand, Chinese subject RCs are also more costly in terms of having to resolve a longer linear relationship between the filler and the gap [compare the distance between the gap and the filler (i.e., head noun) "mouse" in (1) and (2) vs. the distance between the gap and the head noun "chicken" in (3) and (4)], and in terms of deviating from the canonical SVO word order (Bever, 1970; Gibson, 1998, 2000; Diessel and Tomasello, 2005). What makes Chinese RCs intriguing is that, unlike English, subject prominence or structural influences, and linear influences such as similarity to canonical SVO word order and shorter linear distance between filler and gap, are no longer confounded to favor subject over object RCs, but work in opposite directions to both favor and disfavor subject RC processing. In Chan et al.

(2011) and subsequently in Kidd et al. (2015), we argued that the processing and acquisition of Chinese RCs bear on the theoretical themes of competition and variation (MacWhinney, 2005, 2012).

## Chinese Relative Clause Processing and Cross-Linguistic Influences in Multilingual Acquisition

We now turn to discussing why the processing of Chinese RCs is interesting in a multilingual acquisition context. We focus on cross-linguistic influence, where structural overlaps between languages have been identified as a pre-condition for transfer (Hulk and Müller, 2000; see section Current Study and Hypotheses for further elaborations). First, since Cantonese and Mandarin are typologically close, their RCs overlap both structurally and functionally. For instance, the Cantonese subject RC in (1) and the Mandarin subject RC in (2) are highly similar. Likewise, the Cantonese object RC in (3) and the Mandarin object RC in (4) are highly similar. It is therefore reasonable to expect that the structural and functional overlaps between Cantonese and Mandarin RCs could provide a transparent basis for positive transfer between these two Chinese languages when individuals learn these two languages in multilingual acquisition. By contrast, the second point relates to vulnerability to negative cross-linguistic influence in multilinguals, and this requires us to first highlight an important difference between Cantonese and Mandarin RCs. As mentioned, Cantonese object RCs and Mandarin object RCs are highly similar [compare (3) and (4)]; however, there is also an important difference between them in terms of degree of overlap with SVO main clauses. One important structural feature unique to Cantonese object classifier RCs [see (3)] is that they share an identical surface structure with a SVO main clause, and as such instantiate a complete structural overlap with SVO transitive main clauses. Compare the object classifier



RC in (3) repeated below as (5) and a Cantonese SVO main clause in (6).

#### Cantonese object classifier RC

- (5) [RC 老鼠 錫 \_\_\_\_ j] [head noun 嗰 隻 公雞]  
 lou5syu2 sek3 go2 zek3 gung1gai1  
 mouse kiss that CL chicken  
 "The chicken that the mouse kisses"

#### Cantonese transitive SVO main clause

- (6) [MC 老鼠 錫 嗰 隻 公雞]  
 lou5syu2 sek3 go2 zek3 gung1gai1  
 mouse kiss that CL chicken  
 "The mouse kisses the chicken"

Interestingly, a recent study by Lau (2016a) elicited native Cantonese adult speakers' production of object classifier RCs like (5) in one condition and their production of transitive SVO main clauses like (6) which were identical in surface form in another condition, and the acoustic analyses found no prosodic differences between examples like (5) vs. (6). The results suggested that adult native speakers of Hong Kong Cantonese do not use prosody to disambiguate surface identity in syntax between object classifier RCs and transitive main clauses. Note that this characteristic of surface identity is unique to Cantonese object classifier RCs, but not Mandarin object RCs, because Mandarin object RCs [see (7)] only resemble but are not identical in surface structure with SVO main clauses, due to the presence of the relative marker *de*. Compare the Mandarin object RC repeated in (7) and a Mandarin SVO main clause in (8).

#### Mandarin object RC

- (7) [RC 老鼠 親 \_\_\_\_ j] 的 [head noun 公雞]  
 laoshu qing de gongji  
 mouse kiss de chicken  
 "The chicken that the mouse kisses"

#### Mandarin transitive SVO main clause

- (8) [MC 老鼠 親 公雞]  
 laoshu qing gongji  
 mouse kiss chicken  
 "The mouse kisses the chicken"

This surface identity between object classifier RCs and SVO main clauses in Cantonese presents advantages and challenges in acquisition and processing. Merit-wise, in Chan et al. (2011), we argued that Cantonese object classifier RCs allow for and could be facilitated by an internally headed RC analysis. Specifically, object classifier RCs like (5) can be analyzed as internally headed RCs as (9):

- (9) [NP/S lou5syu2 sek3 go2 zek3 gung1gai1]  
 mouse kiss that CL chicken.  
 "The chicken that the mouse kisses"

Under the internally headed RC analysis, example (9) has the internal structure of a SVO clause, but behaves as a noun phrase (NP) in terms of its external syntax. The internally headed RC analysis is represented by the notation NP/S in (9) above,

indicating that a constituent has externally the syntax of a NP but internally that of a clause (S). Here the internal structure is a SVO main clause, with the object, which is also the head noun, *in situ*. Hence the head "chicken" is internal to the RC. Internally headed RCs do not involve gaps or extraction, are structurally simpler, and therefore may be easier to process than externally headed RCs (see Jeon and Kim, 2007 for supportive evidence from Korean). This internally headed analysis is only possible for Cantonese object classifier RCs because it is only in this case where there is complete surface identity with simple main clauses and therefore ambiguity of analysis. Examples like (5), as such, are structurally ambiguous as they can be analyzed as head-final RCs (5) or internally headed RCs (9).

Moreover, Cantonese learners could make use of simple transitives to bootstrap onto Cantonese object RCs, especially of the classifier type, in production. On the other hand, we also acknowledged that their surface identity with SVO main clauses could cause problems in comprehension, notably by leading Cantonese object classifier RCs to be mis-parsed as SVO transitive main clauses (Lau, 2016b), due to structural ambiguity. The potential to be misled due to competing analyses in sentence parsing could become more complicated for multilingual children acquiring Cantonese RCs under heavy influence from English, especially when there are additional competing constructions due to structural overlaps between the children's languages. This brings us back to the second point about vulnerability to negative cross-linguistic influence in multilinguals. Specifically, parsing of Cantonese object classifier RCs could be especially challenging for these multilingual children, because the Cantonese object RCs not only overlap with SVO in Cantonese but also SVO transitive clauses and head-initial subject RCs in English. Compare (5) and (6) repeated below as (10) and (11), alongside the English transitive SVO main clause in (12) and the English subject RC in (13).

#### Cantonese object classifier RC (head-final)

- (10) [RC 老鼠 錫 \_\_\_\_ j] [head noun 嗰 隻 公雞]  
 lou5syu2 sek3 go2 zek3 gung1gai1  
 mouse kiss that CL chicken  
 "The chicken that the mouse kisses"

#### Cantonese transitive SVO main clause

- (11) [MC 老鼠 錫 嗰 隻 公雞]  
 lou5syu2 sek3 go2 zek3 gung1gai1  
 mouse kiss that CL chicken  
 "The mouse kisses the chicken"

#### English transitive SVO main clause

- (12) The mouse kisses the chicken

#### English subject RC (head-initial)

- (13) [head noun The mouse;] [RC that \_\_\_\_ j kisses the chicken]

Overlap with English head-initial subject RCs (also SVO) may encourage a head-initial analysis here. In particular, when Cantonese (head-final) object classifier RCs lack an overt relative marker introducing the head noun of the RC, head noun

assignment could be especially vulnerable to negative cross-linguistic influence from English.

In fact, cross-linguistic influences have been observed in our previous work on simultaneous Cantonese-English bilingual children. Yip and Matthews (2007a) analyzed naturalistic speech production and found that object relative clauses (the classifier type, such as (3) but often with an inanimate head noun and an animate subject NP in the RC) emerged earlier than or simultaneously with subject relative clauses [such as (1)] in the bilingual children's Cantonese; while in their English, Cantonese-based prenominal relatives emerged first, with object relatives (e.g., "Where's [<sub>NP</sub> you buy that one]" meaning "where's the one you bought" (example 15 from Yip and Matthews, 2007b) followed by subject relatives (e.g., "I want [<sub>NP</sub> have ear that one]" meaning "I want the one that has ears" (example 20 from Yip and Matthews, 2007a). On the other hand, in a comprehension experiment, Kidd et al. (2015) found that their bilingual children made more head noun errors than the monolinguals when comprehending Cantonese object RCs that are consistent with an English head-initial analysis, erroneously choosing the subject of the RC [the first noun of the complex noun phrase, i.e., the "mouse" in (10)] rather than the "chicken" in (10) as the head noun.

Looking broader beyond Chinese and English in the context of the current literature, we highlight the following observations. First, descriptions of cross-linguistic influence in trilingual acquisition have largely featured adult learners and English and European languages (e.g., Cenoz and Jessner, 2000; Cenoz et al., 2001). Studies featuring cross-linguistic interactions in trilingual children exist, but many of which are case studies featuring a few children (Hoffmann, 1985; Helot, 1988; Li, 2006 *inter alia.*). Experimental studies testing a group of trilingual children have been relatively few. Regarding cross-linguistic influences in trilingual children, the broad trends of investigation have been on reporting the observed code-switching and mixing patterns between languages (e.g., Stavans and Swisher, 2006; Edwards and Dewaele, 2007; Hoffmann and Stavans, 2007; Stavans and Muchnik, 2008) and how the prior languages affect the acquisition of a third language (e.g., Oksaar, 1977; Hoffmann, 1985; Flynn et al., 2004; Anastassiou and Andreou, 2014). For instance, Oksaar (1977) identified negative transfer of semantics of verbs from the two L1s Estonian and Swedish of a child to his L3 German. However, we know relatively little about how the latter acquired languages affect the prior acquired languages (so called "reverse" transfer) from the current literature on trilingual children. A notable exception is Kazzazi (2011), which approached cross-linguistic influence in 2 trilingual children from a cognitive perspective. This study found that the post-modifying order in the non-dominant language Farsi was transferred to the other two languages (German and English) because this order manifests the general cognitive tendency toward iconicity and transparency. Thus, far there has been very little research on childhood trilingualism which approaches the issue of cross-linguistic influence from the theoretical perspective of structural overlaps between languages. On the other hand, cross-linguistic transfer due to structural overlaps has been more intensively

studied in the bilingualism literature (see Serratrice, 2013 for a review).

## Current Study and Hypotheses

As a follow-up to our previous works (Yip and Matthews, 2000, 2007a; Kidd et al., 2015), we extend our work on cross-linguistic influences by examining a new group of multilingual children. Unlike our previous work that investigated simultaneous Cantonese-English bilingual children in Hong Kong (Yip and Matthews, 2000, 2007a) and in Australia (Kidd et al., 2015), we target a group of Cantonese-English-Mandarin trilingual children that is unique and relevant to a significant number of children in Hong Kong studying in international schools/curriculums with an English immersion environment from an early age. These children acquire Cantonese as their family and first language, and also acquire English and Mandarin as second and third languages at school. Although English is not the community language of Hong Kong, these children's Chinese is under heavy influence from English because they are educated in an English immersion environment. Specifically, we tested how Cantonese-English-Mandarin trilingual children's comprehension of subject and object RCs was influenced by the structural overlaps between the three languages when the two Chinese languages are acquired under different exposure conditions. These patterns of overlaps and differences may raise new possibilities for interactions between the three developing linguistic systems in trilingual children.

The current study draws reference to a number of theoretical perspectives in bilingual and multilingual acquisition and the kinds of transfer these perspectives predict. In particular, we draw reference to Hulk and Müller's specific hypothesis related to cross-linguistic influence in childhood bilingualism research (Hulk and Müller, 2000; Müller and Hulk, 2001). In addition, we consider several factors that have been proposed to drive the directionality of cross-linguistic influences, namely, typological distance, psychotypology, and language dominance. Hulk and Müller's hypothesis and these factors will be introduced briefly below, which will contribute to the formulation of our hypotheses specific to the current study.

In Hulk and Müller's hypothesis, one necessary condition for cross-linguistic influence to occur is partial structural overlap between the two languages regarding the structure of interest. Their original hypothesis defined the structural overlap condition as such: "syntactic cross-linguistic influence occurs only if language A has a construction which may seem to allow more than one syntactic analysis and, at the same time, language B contains evidence for one of these two possible analyses. In other words there has to be a certain overlap of the two systems at the surface level" (Hulk and Müller, 2000, p. 228–229). According to this hypothesis, if a structure in language A is potentially ambiguous between more than one analysis, and that language B allows only one of the analysis, there will be unidirectional influence from language B to language A in that the overlapping analysis would be adopted by the bilinguals more often than by the monolinguals. Another potential factor affecting directionality of cross-linguistic influence is typological distance (or linguistic distance). It has been proposed to be a

major factor in the choice of the source language regarding cross-linguistic influence in multilingual language acquisition (Cenoz, 2001). This perspective is supported by the observation that speakers tend to transfer more vocabulary items and structures from the language that is typologically closer to the target language. A related notion is the concept of psychotypology by Kellerman (1983), that is, the language that is “perceived” as typologically closer. The role of psychotypology has been demonstrated in the literature. For instance, learners of English and French whose first language is a non-Indo-European language would tend to transfer vocabulary and structures from other Indo-European languages they know rather than from their L1 (Ahukanna et al., 1981; Ringbom, 1987; Bartelt, 1989). In addition, language dominance is another factor that can predict cross-linguistic influence: the source language tends to be the more dominant language (Yip and Matthews, 2000, 2007b).

We have two hypotheses focusing on the two Chinese languages for the current study. First we hypothesize that these trilingual children would experience facilitation in comprehending RCs in their third language Mandarin even with limited exposure, due to positive influence from their first language Cantonese. In particular, we expect that positive transfer from Cantonese to Mandarin allows the trilingual children to comprehend Mandarin RCs above the level that would be expected based on their limited input (as reflected by their weak vocabulary knowledge in Mandarin). Here we take vocabulary score as a proxy variable for a child's language-specific experience, and therefore expect that the trilinguals' Mandarin vocabulary scores would be significantly lower than their age matched monolingual Mandarin peers. However, by contrast, we expect that the trilinguals would not score as much lower than their monolingual age peers in their Mandarin RC comprehension performance as in their vocabulary scores, and they might even perform comparable to their monolingual age peers. The first hypothesis is motivated by the typological close proximity between Cantonese and Mandarin, and their similar RC structures in particular [compare (1) and (2)], coupled with the fact that Cantonese is the more dominant language while Mandarin is the weaker language for the trilingual children under investigation.

Second, we hypothesize that these trilingual children would experience more difficulty in comprehending Cantonese object classifier RCs relative to their monolingual peers, especially in head noun assignment, due to negative influence from English and intensive exposure to English. We therefore expect that the trilinguals would make significantly more head noun errors than their monolingual peers when comprehending Cantonese object classifier RCs, with the error pattern consistent with an English-based head-initial analysis. This hypothesis is motivated by the consideration that Cantonese object classifier RCs are potentially ambiguous between more than one analysis as described in section Trilingual vs. Monolingual Mandarin above, and these Cantonese object classifier RCs overlap with subject RCs in English when the two languages are in contact in a multilingual child, while English RCs clearly allow only a head-initial analysis. As such, transfer from English to Cantonese is possible based on Hulk and Müller's hypothesis.

## METHODS

### Participants

Sixty-nine ( $N = 69$ ) children participated. Twenty-one ( $N = 21$ , 10 females) Cantonese(L1)-English(L2)-Mandarin(L3) trilingual children were recruited from an international English-immersion elementary school in Hong Kong. Twenty-four ( $N = 24$ , 11 females) predominantly monolingual Cantonese-speaking children in Hong Kong, and 24 ( $N = 24$ , 11 females) monolingual L1 Mandarin children in China, served as comparison groups for the two Chinese languages. The predominantly monolingual Cantonese children were born in Hong Kong, spoke Cantonese at home, and the primary language of instruction at school is Cantonese. The trilingual group was aged between 5;4 and 6;1 ( $M_{\text{age}} = 5;8$ ,  $SD = 0;2$ ). The comparison groups were matched by age for both Cantonese and Mandarin: the monolingual Cantonese group was aged between 5;4 and 6;4 ( $M_{\text{age}} = 5;11$ ,  $SD = 0;3$ ) and the monolingual Mandarin group was aged between 5;9 and 6;5 ( $M_{\text{age}} = 5;11$ ,  $SD = 0;2$ ). Our trilingual English data showed the subject over object RC advantage well-attested in English, so we did not test a monolingual English comparison group. **Table 1** summarizes the participant information.

The trilingual children come from Hong Kong middle class families with both parents being native speakers of Cantonese. They have been exposed to Cantonese in the family and community from birth. These children became regularly and intensively exposed to English when they entered kindergarten around the age of 3. At the time of testing, they were attending an international English-immersion primary school five and a half hours a day and 5 days a week, during which they also received regular but far less extensive exposure to Mandarin as a foreign language for 200 min per week. The children reported speaking both Cantonese and English at home.

### Trilingual Children's Language Proficiency

The Cantonese Receptive Vocabulary Test (CRVT; Cheung et al., 1997) was used to assess the children's receptive Cantonese vocabulary knowledge. This standardized test provides norms based on monolingual Cantonese children in Hong Kong aged 2;0–6;0, giving some objective measure of the children's proficiency in Cantonese<sup>1</sup>. For Cantonese, the trilinguals scored

**TABLE 1 |** Subject information.

|           | Trilingual Cantonese-Mandarin-English | Monolingual Cantonese | Monolingual Mandarin |
|-----------|---------------------------------------|-----------------------|----------------------|
| N         | 21                                    | 24                    | 24                   |
| Age Range | 5;4–6;1                               | 5;4–6;4               | 5;9–6;5              |
| Mean age  | 5;8 ( $SD = 0;2$ )                    | 5;11 ( $SD = 0;3$ )   | 5;11 ( $SD = 0;2$ )  |

<sup>1</sup>We did not use standardized language assessments such as Reynell Developmental Language Scales (Reynell and Huntley, 1987) and HKCOLAS (T'sou et al., 2006). Although they offer more comprehensive information including morphosyntax, they take much longer to run than could be accommodated by the school.

**TABLE 2 |** Vocabulary Scores of the Trilingual Group (chronological age:  $M = 5;8$ ,  $SD = 0;2$ ).

|                   | Cantonese vocabulary |         | English vocabulary |         | Mandarin vocabulary                                        |       |
|-------------------|----------------------|---------|--------------------|---------|------------------------------------------------------------|-------|
|                   | <i>M (SD)</i>        | Range   | <i>M (SD)</i>      | Range   | <i>M (SD)</i>                                              | Range |
| <b>TRILINGUAL</b> |                      |         |                    |         |                                                            |       |
| Raw score         | 60 (3)               | 53–64   | 52.5 (6.2)         | 39–63   | 87.4 (9.99)                                                | 67–99 |
| Percentage (%)    | 92 (5)               | 82–98   | 31 (4)             | 23–38   | 82 (9)                                                     | 63–93 |
| Age equivalent    | 5;8 (0;4)            | 5;0–6;1 | 5;2 (0;8)          | 3;8–6;5 | Scores lower than 3-year-old monolingual Mandarin children |       |

on average 60 out of a total of 65 items in the CRVT correct. The majority of the trilingual children scored comparably to their monolingual age peers in the normative sample of the test (age equivalent according to CRVT:  $M_{age} = 5;8$ ,  $SD = 0;4$ ,  $Range = 5;0–6;1$ ), with only 3 children scoring 1  $SD$  or more below mean. These 3 children were still included as their data do not change the results, and their inclusion increased the power of the analyses. The British Picture Vocabulary Scale 2 (BPVS2; Dunn et al., 1997) was used to assess the children's receptive English vocabulary knowledge. This standardized test provides norms based on monolingual English children in UK aged 3–15, giving some objective measure of the children's proficiency in English. For English, the trilinguals scored on average 52.5 out of a total of 168 items correct (there were more items in BPVS than CRVT as the former can be used for older children), and their performance is more variable (age equivalent according to BPVS:  $M_{age} = 5;2$ ,  $SD = 0;8$ ,  $Range = 3;8–6;5$ ). For Mandarin, we used a receptive vocabulary test we have developed (Chan et al., 2014) that assesses comprehension of 106 words from 14 semantic categories that are chosen based on the early vocabulary inventory of Mandarin-speaking children in Beijing (Hao et al., 2008). As expected, the trilinguals scored significantly lower than the monolingual age-matched comparison group in their Mandarin vocabulary scores [ $t_{(22)} = -5.9$ ,  $p < 0.000$ ,  $d = 1.80$ ]. In fact, these 5- to 6-year-old trilinguals' performed even worse than the 3-year-old monolingual Mandarin group ( $N = 49$ ; aged 2;11–3;05) in the normed sample of the Mandarin receptive vocabulary test (percentage accuracy: trilinguals:  $M = 0.82$ ,  $SD = 0.09$ ; monolinguals:  $M = 0.93$ ,  $SD = 0.036$ ). **Tables 2, 3** show the children's performance on the vocabulary tests.

## Materials and Procedure

All children were tested individually by a female experimenter in a quiet room in their school. All children were tested by a native speaker of the respective language. The trilingual children were tested in three sessions, one for each language (vocabulary test first, and then RC test), with the sequence of the languages tested counterbalanced between children.

## Test of Vocabulary Knowledge

Test administration followed the standardized test instructions for the CRVT (Cheung et al., 1997), the BPVS2 (Dunn et al., 1997), and the Mandarin receptive vocabulary test (Chan et al., 2014). For all the three vocabulary tests, children were presented with 4 pictures showing the target word and 3 distractor pictures,

**TABLE 3 |** Vocabulary Scores of the Monolingual Mandarin Comparison Group (chronological age:  $M = 5;11$ ,  $SD = 0;2$ ).

|                    | Mandarin vocabulary |        |
|--------------------|---------------------|--------|
|                    | <i>M(SD)</i>        | Range  |
| <b>MONOLINGUAL</b> |                     |        |
| Raw score          | 101.3 (3.3)         | 93–106 |
| Percentage (%)     | 96 (3)              | 88–100 |

and were asked to point to the picture that matched a spoken word.

We did not test the monolingual Cantonese children with CRVT, because their CRVT scores were not needed given the following reasons. First, before being confirmed to be able to take part in the study, the monolinguals had been screened by a speech therapist to ensure that they did not present any noticeable speech and language delays in their L1 Cantonese at the time of testing. Second, we intended to match the trilinguals and the Cantonese monolinguals only on their chronological age but not on language proficiency, as the time required for running a full standardized language assessment could not be accommodated by the school. In addition, matching the two groups only on the basis of receptive vocabulary measures to claim for language-matched status is not unproblematic. Third, obtaining the monolinguals' CRVT scores or not would not affect the main pattern of the current findings and their interpretations (see section Discussion for further elaborations).

## Test of RC Comprehension

We used the sentence interpretation pointing method and its materials established in Kidd et al. (2015), described briefly below (see Kidd et al., 2015 for details). Children were shown pairs of pictures on a computer screen. Within each pair, both pictures showed the same causative event between two animals and differed only in which animal was the agent and the patient of the action e.g., one picture showed a cat feeding a duck and the other a duck feeding a cat, see **Figure 3**). Children heard test sentences such as *Where's the duck that is feeding the cat? Find it!* (subject-relative) or *Where's the duck that the cat is feeding? Find it!* (object-relative), and were asked to point to the animal described by the experimenter. Each child received 8 Subject(Agent)- RC test

sentences and 8 Object(Patient)- RC test sentences as stimuli for the language being tested, with length and animacy controlled. **Table 4** shows examples of the sentence stimuli in the three languages.

### Data Coding

Children's responses were coded into four categories: (i) Correct, (ii) Head error: when children pointed to the correct picture but the incorrect animal (e.g., pointing to the cat in the correct picture for the test sentence *Where's the duck that the cat is feeding?*) (iii) Reversal error: when children pointed to the correct token of the head referent in the incorrect picture (e.g., pointing to the picture where the duck is the agent for the test sentence *Where's the duck that the cat is feeding?*), and (iv) Other error: when children pointed to the incorrect animal in the incorrect picture (e.g., pointing to the cat in the incorrect picture for the test sentence *Where's the duck that the cat is feeding?*). The first author coded all the children's responses. One research assistant from each language coded at least 20 percent of the data (at least 10 children from each language) for inter-rater reliability. Inter-rater reliability was close to 100% agreement in all cases.

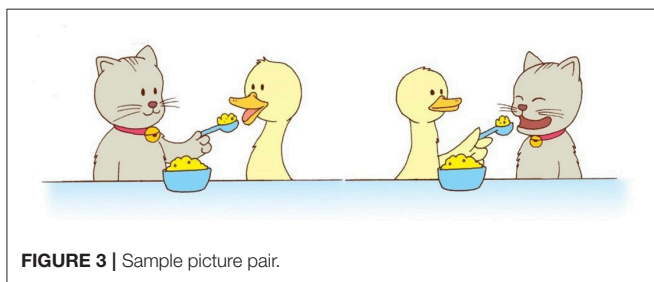


**FIGURE 3** | Sample picture pair.

## RESULTS

**Figure 4** shows the trilingual and monolingual groups' performance on the subject vs. object RCs in Cantonese and Mandarin, and the trilinguals' performance on the English subject- and object- RCs. The trilingual children comprehended subject RCs better than the object RCs for all the three languages (Cantonese:  $M_{\text{subjRC}} = 0.60$ ,  $M_{\text{objRC}} = 0.29$ ; Mandarin:  $M_{\text{subjRC}} = 0.62$ ,  $M_{\text{objRC}} = 0.34$ ; English:  $M_{\text{subjRC}} = 0.91$ ,  $M_{\text{objRC}} = 0.30$ ). The monolingual Mandarin children also comprehended subject RCs better than the object RCs ( $M_{\text{subjRC}} = 0.59$ ,  $M_{\text{objRC}} = 0.38$ ). In contrast, the monolingual Cantonese children found object RCs easier to comprehend than subject RCs ( $M_{\text{subjRC}} = 0.46$ ,  $M_{\text{objRC}} = 0.58$ ). In the following sections, we used the R package lme4 (Bates and Maechler, 2010) in R (version 3.3.1, R Core Development Team, 2016) to fit generalized linear mixed models (Jaeger, 2008). The final model was chosen based on significance of fixed effects and random effects. Only significant terms were included.

### Overall Analysis

The monolingual and trilingual children's correct responses in Cantonese and Mandarin were analyzed first. The data were analyzed using Generalized Linear Mixed Models (GLMM). This analysis was to test whether there is a significant interaction between Group and Extraction. The fixed effects were Group (trilingual vs. monolingual), Extraction (subject vs. object) and their interaction. Random effects for participants were included to model variation among participants (random intercepts), and by-participant random slopes were also included if significant. Random slopes for the variable of Extraction contributed to model fit significantly and were included in the model. There was a significant Group  $\times$  Extraction interaction ( $\beta = 2.1$ ,  $z = 2.32$ ,  $p = 0.02$ ). This interaction was therefore further scrutinized by

**TABLE 4** | Examples of test sentences in the three languages.

| Sentence type | Example                                          |           |         |        |    |       |      |
|---------------|--------------------------------------------------|-----------|---------|--------|----|-------|------|
| Sub-Eng       | Where's the cat that is feeding the duck?        |           |         |        |    |       |      |
| Sub-Can       | 錫緊                                               | 公雞        | 嗰隻      | 老鼠     | 喺  | 邊度    | 呀?   |
|               | kiss-PROG                                        | chicken   | that-CL | mouse  | is | where | SFP? |
|               | "Where's the mouse that is kissing the chicken?" |           |         |        |    |       |      |
| Sub-Man       | 抱                                                | 小豬        | 的       | 小狗     | 在  | 哪     | ?    |
|               | hug                                              | piggy     | de      | doggy  | is | where | ?    |
|               | "Where's the dog that is hugging the pig?"       |           |         |        |    |       |      |
| Obj-Eng       | Where's the horse that the pig is hugging?       |           |         |        |    |       |      |
| Obj-Can       | 羊仔                                               | 推緊        | 嗰隻      | 兔仔     | 喺  | 邊度    | 呀?   |
|               | sheep                                            | push-PROG | that-CL | rabbit | is | where | SFP? |
|               | "Where's the rabbit that the sheep is pushing?"  |           |         |        |    |       |      |
| Obj-Man       | 白馬                                               | 餵         | 的       | 老虎     | 在  | 哪     | ?    |
|               | white.horse                                      | feed      | de      | tiger  | is | where | ?    |
|               | "Where's the tiger that the horse is feeding?"   |           |         |        |    |       |      |

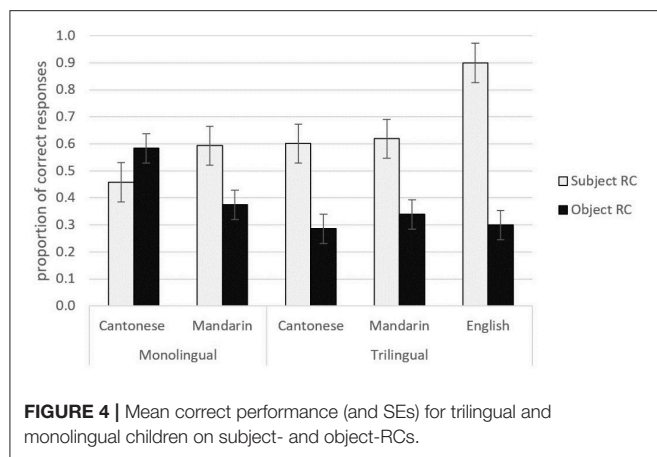


**FIGURE 4 |** Mean correct performance (and SEs) for trilingual and monolingual children on subject- and object-RCs.

**TABLE 5 |** Significant terms in final model for analysis of RC Comprehension in Trilingual vs. Monolingual Cantonese.

|                    | $\beta$ | SE    | z      | p                       |
|--------------------|---------|-------|--------|-------------------------|
| Intercept          | 0.392   | 0.251 | 1.562  | 0.118                   |
| Extraction         | −0.600  | 0.225 | −2.665 | 0.008 <sup>***a</sup>   |
| Group              | −1.30   | 0.318 | −4.086 | <0.001 <sup>****a</sup> |
| Group × Extraction | 2.499   | 0.271 | 9.222  | <0.001 <sup>****a</sup> |

<sup>a</sup>log likelihood = −829.7, Number of observations = 1,391, <sup>\*\*\*</sup> $p < 0.001$ , <sup>\*\*</sup> $p < 0.01$ , <sup>\*</sup> $p < 0.05$ .

analyzing the trilingual vs. monolingual Cantonese groups and the trilingual vs. monolingual Mandarin groups separately using the same analysis strategy.

## Trilingual vs. Monolingual Cantonese

Similar to overall analysis in section Overall Analysis, we used GLMM. The analysis was to test whether Group (trilingual vs. monolingual Cantonese), Extraction (subject vs. object) and their interaction significantly contributed to the responses. The fixed effects were Group (trilingual vs. monolingual), Extraction (subject vs. object) and their interaction. Random effects for participants were included to model variation among participants (random intercepts). Random slopes for the variable of Group contributed to model fit significantly and were included in the model. Models were compared with and without random effects (random intercepts or slopes) by likelihood ratio tests to test the significance of them. The final model only included significant random effects. The significant effects for the final model are shown in **Table 5**. There were significant effects of Extraction and Group, and a significant Group × Extraction interaction. *Post-hoc* analyses that analyzed each extraction type separately showed that the group difference lay crucially in object but not subject RCs. Specifically, the trilinguals comprehended the Cantonese object RC sentences significantly worse than the monolinguals ( $\beta = -2.6$ ,  $z = -2.48$ ,  $p = 0.01$ ). When comprehending Cantonese subject RC sentences, the trilinguals performed slightly better than the monolinguals, though the difference was not significant ( $\beta = 0.87$ ,  $z = 1.6$ ,  $p = 0.12$ , *n.s.*).

**TABLE 6 |** Significant terms in final model for analysis of RC comprehension in Trilingual vs. Monolingual Mandarin.

|                   | $\beta$ | SE    | z      | p                   |
|-------------------|---------|-------|--------|---------------------|
| Intercept         | −0.904  | 0.492 | −1.838 | 0.066               |
| Extraction        | 1.422   | 0.657 | 2.166  | 0.030 <sup>ab</sup> |
| Group             | −0.355  | 0.686 | −0.518 | 0.605               |
| Group: extraction | 0.974   | 0.924 | 1.054  | 0.292               |

<sup>b</sup>log likelihood = −728.1, Number of observations = 1,391, <sup>\*\*\*</sup> $p < 0.001$ , <sup>\*\*</sup> $p < 0.01$ , <sup>\*</sup> $p < 0.05$ .

## Trilingual vs. Monolingual Mandarin

Similar analysis was conducted to test whether Group (trilingual vs. monolingual Mandarin), Extraction (subject vs. object) and their interaction significantly contributed to the responses. In the GLMM we fit, the fixed effects were Group (trilingual vs. monolingual), Extraction (subject vs. object) and their interaction. Random effects for participants were included to model variation among participants (random intercepts). Random slopes for the variable of Extraction contributed to model fit significantly and were included in the model. The significant effects for the final model are shown in **Table 6**. The only significant effect was that of Extraction, indicating that children comprehended subject RCs better than object RCs in general. Crucially, there was no significant effect of Group and Group did not interact with Extraction, showing that the trilinguals and monolinguals were performing similarly when comprehending Mandarin RCs. This result is interesting because the trilinguals showed similar performance to their age-matched monolingual peers in Mandarin, despite Mandarin being their third and weaker language due to limited exposure (recall that the trilinguals scored significantly lower than the monolinguals in their receptive Mandarin vocabulary, and these trilinguals' receptive Mandarin vocabulary scores were even lower than the 3-year-old monolingual Mandarin group in the normed sample of the vocabulary test). In addition, as **Figure 4** shows, the trilinguals displayed strikingly similar performance profiles when comprehending subject and object RCs in their L1 Cantonese and L3 Mandarin, suggesting that positive transfer from Cantonese to Mandarin is implicated. We will return to this point in section Positive Transfer from L1 Cantonese to L3 Mandarin.

Next, we further analyzed data from each group (monolingual Cantonese children, monolingual Mandarin children, and trilingual children) separately.

## Monolingual Cantonese

For monolingual Cantonese children, we tested whether Extraction was significant. In the GLMM, the fixed effect was Extraction (subject vs. object), and the random effects for participants were included to model variation among participants (random intercepts). Random slopes for the variable of Extraction contributed to model fit significantly and were included in the model. Analyses of the monolingual Cantonese data revealed a non-significant effect for Extraction ( $\beta = -0.83$ ,  $z = -0.98$ ,  $p = 0.3$ , *n.s.*), indicating that although

the monolingual Cantonese children found object RCs easier to comprehend than subject RCs as shown in **Figure 4**, the difference was not significant. This slight object advantage is consistent with past comprehension studies on monolingual Cantonese-speaking children's processing of classifier RCs using the same pointing method (Chan et al., 2011; Kidd et al., 2015) as well as using a referent selection eye-tracking task to yield online processing data (Chan et al., 2017).

## Monolingual Mandarin

For monolingual Mandarin children, we tested whether Extraction and Mandarin vocabulary scores were significant. Extraction and Mandarin vocabulary scores were entered as fixed effects. Random slopes for Extraction contributed significantly and were included in the model. There was a significant effect for Extraction ( $\beta = 3.16$ ,  $z = 1.98$ ,  $p = 0.048$ ), meaning that the monolingual Mandarin children comprehended subject RCs significantly better than the object RCs, as shown in **Figure 4**. This subject advantage is consistent with recent experimental findings on monolingual Mandarin-speaking children's processing of RCs (Hsu, 2014; Hu et al., 2015a,b). There was no significant effect for Mandarin vocabulary, likely due to our monolingual children scoring close-to-ceiling in the vocabulary test.

## Trilingual Data

We tested whether Extraction was significant for the trilingual children, and whether Mandarin, Cantonese and English vocabulary scores significantly predicted these trilinguals' RC performance. Analyses of the trilingual data revealed a significant effect for Extraction ( $\beta = 2.36$ ,  $z = 4.84$ ,  $p < 0.001$ ). Random intercepts and slopes of the variables Extraction, Mandarin Vocabulary and English Vocabulary were significant and were included. There was a marginally significant effect for Mandarin vocabulary as a predictor of the trilinguals' Mandarin RC performance:  $\chi^2(1) = 3.09$ ,  $p = 0.079$  and a marginal significance for Cantonese vocabulary as a predictor of the trilinguals' Cantonese RC performance: Cantonese  $\chi^2(1) = 3.62$ ,  $p = 0.057$ ), showing that, unsurprisingly, children's RC comprehension in a language improved as their vocabulary scores in the target language increased. There was no significant or a marginally significant effect for English Vocabulary though. We then examined the trilinguals' performance in each language. Random slopes for Extraction contributed significantly and were included in each model for each language. There was a significant effect for Extraction in each language, indicating a significant advantage for subject over object RCs in the trilinguals' L1 Cantonese as well as their L2 English and L3 Mandarin, as shown in **Figure 4** (Cantonese:  $\beta = 2.53$ ,  $z = 2.39$ ,  $p = 0.017$ ; Mandarin:  $\beta = 1.86$ ,  $z = 2.66$ ,  $p = 0.0078$ ; English:  $\beta = 4.87$ ,  $z = 5.19$ ,  $p < 0.001$ ).

## Positive Transfer from L1 Cantonese to L3 Mandarin

In order to further address the likelihood of positive transfer from L1 Cantonese to L3 Mandarin in these trilingual children, we carried out the following analyses. First, we tested whether

there were any differences in the proportion of correct responses between the trilinguals' Cantonese vs. Mandarin by fitting mixed effects logistic regression models. We fit two models for a comparison to test a fixed effect of language. The first model only modeled by-participant random intercepts and treats the trilingual's Cantonese and Mandarin as having the same proportion of correct responses. The second model added the fixed effect of language, treating the trilingual's Cantonese and Mandarin as having different proportion of correct responses. A likelihood ratio test was employed to compare the goodness of fit of two models, and test whether there are significant differences between the two models. If there are significant differences, it means that the trilingual's Cantonese and Mandarin have different proportion of correct responses. A likelihood ratio test showed that there was no significant difference between the two models [ $\chi^2(1) = 0.93$ ,  $p = 0.33$ ], suggesting that the trilinguals' performance in comprehending Cantonese vs. Mandarin RC sentences was similar. Crucially, we also applied the same procedure to test the proportion of correct responses between the trilinguals' Cantonese vs. English in the RC comprehension tasks, but there were significant differences [ $\chi^2(1) = 18.33$ ,  $p < 0.001$ ], meaning that the trilingual's performance in comprehending Cantonese vs. English RC sentences was different. As **Figure 4** shows, there were more correct responses in English than in Cantonese, especially in the subject RC condition. Similarly, there were significant differences in the proportion of correct responses between the trilinguals' Mandarin vs. English in the RC comprehension tasks [ $\chi^2(1) = 11.7$ ,  $p < 0.001$ ].

Second, to investigate whether the trilinguals' L1 Cantonese RC performance can significantly predict L3 Mandarin RC performance, we fitted a generalized linear mixed effects model, where the response was their binary response in comprehending Mandarin RCs, and the fixed effect was their response in comprehending Cantonese RCs. By-participant random slopes were also included due to significance. Results showed that the trilinguals' Cantonese RC correct performance did significantly contribute to their Mandarin RC correct performance ( $\beta = 2.89$ ,  $z = 7.3$ ,  $p < 0.001$ ). Third, additionally, a linear model was fitted with the trilinguals' Mandarin RC scores (the sum of a child's subject and object RC correct responses in the Mandarin RC task) as the responses and their Cantonese RC scores (the sum of a child's subject and object RC correct responses in the Cantonese RC task) as a covariate. The result showed that the trilinguals' L1 Cantonese RC scores positively predicted their L3 Mandarin RC scores ( $\beta = 0.59$ ,  $t = 2.26$ ,  $p = 0.036$ ), and this effect remained even after adding Mandarin vocabulary as a covariate. Follow up analyses that analyzed each extraction type separately showed that the trilinguals' L1 Cantonese subject RC scores positively predicted their L3 Mandarin subject RC scores and the result was highly significant ( $\beta = 0.78$ ,  $t = 5.81$ ,  $p < 0.001$ ), while the trilinguals' Cantonese object RC scores did not predict their Mandarin object RC scores ( $\beta = 0.32$ ,  $t = 1.54$ ,  $p = 0.14$ , *n.s.*). Importantly, the same analysis strategies were used to examine whether the trilinguals' L1 Cantonese RC scores also predicted their L2 English RC scores, in terms of their combined (subject plus object RC) scores as well as their separate scores

for each extraction type, but their L1 Cantonese RC scores did not predict their L2 English RC scores in all these analyses (all  $p > 0.1$ ).

In addition, we used the same analysis strategies to examine whether the trilinguals' L3 Mandarin RC performance also predicted their L1 Cantonese RC performance, in terms of their combined (subject plus object RC) scores as well as their separate scores for each extraction type. First, we fitted a generalized linear mixed effects model, where the response was their binary response in comprehending Cantonese RCs, and the fixed effect was their response in comprehending Mandarin RCs. By-participant random slopes were also included due to significance. Results showed that the trilinguals' Mandarin RC correct performance did significantly contribute to their Cantonese RC correct performance ( $\beta = 2.75$ ,  $z = 7.86$ ,  $p < 0.001$ ). Moreover, a linear model was fitted with the trilinguals' Cantonese RC scores (the sum of a child's subject and object RC correct responses in the Cantonese RC task) as the responses and their Mandarin RC scores (the sum of a child's subject and object RC correct responses in the Mandarin RC task) as a covariate. The result showed that the trilinguals' L3 Mandarin RC scores significantly positively predicted their L1 Cantonese RC scores ( $\beta = 0.36$ ,  $t = 2.26$ ,  $p = 0.035$ ). Follow up analyses that analyzed each extraction type separately showed that the trilinguals' Mandarin subject RC scores positively predicted their Cantonese subject RC scores and the result was highly significant ( $\beta = 0.82$ ,  $t = 5.81$ ,  $p < 0.001$ ), but the trilinguals' Mandarin object RC scores did not predict their Cantonese object RC scores ( $\beta = 0.34$ ,  $t = 1.54$ ,  $p = 0.14$ , *n.s.*).

To summarize, despite showing similar profiles in comprehending subject RCs better than object RCs in all the three languages (see **Figure 4**), the trilingual children's L1 Cantonese RC scores positively predicted only their L3 Mandarin RC scores but not their L2 English RC scores. In particular, their L1 Cantonese subject RC correct performance strongly and positively predicted their L3 Mandarin subject RC correct performance suggesting positive influence from L1 Cantonese to L3 Mandarin, given the structural parallels between Cantonese and Mandarin RCs as a transparent basis for positive transfer. Interestingly, their L3 Mandarin subject RC correct performance also strongly and positively predicted their L1 Cantonese subject RC correct performance, suggesting that the Cantonese and Mandarin subject RCs share the same representation in these trilinguals. On the other hand, it is also interesting to note that these trilinguals' L1 Cantonese object RC correct performance did not predict their L3 Mandarin object RC correct performance despite the structural overlaps, nor did their L3 Mandarin object RC performance predict their L1 Cantonese object RC performance. This finding is consistent with the idea that children were analyzing the Cantonese object RCs and the Mandarin object RCs differently (which also accords with the linguistic differences between Cantonese and Mandarin object RCs, see section Chinese Relative Clause Processing and Cross-Linguistic Influences in Multilingual Acquisition) and that Cantonese object RCs but not (or to a lesser extent) Mandarin object RCs were subject to cross-linguistic influence from English in these trilinguals.

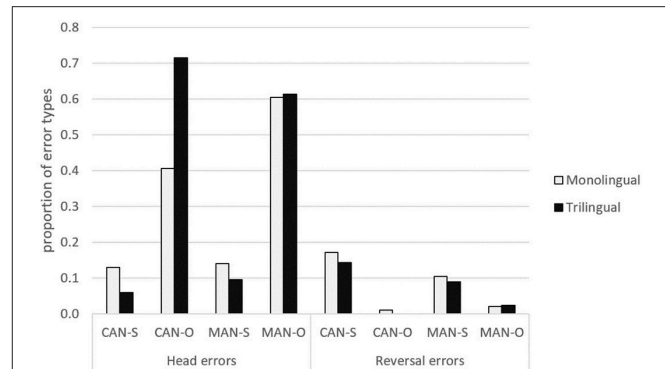


**FIGURE 5 |** Distribution of error types for monolingual and trilingual groups for subject and object RCs.

## Error Analyses

We now turn to analyses of the error responses. Children made three error types: (i) head errors, (ii) reversal errors, and (iii) “other” errors. **Figure 5** shows the monolingual and trilingual children's average error percentage when comprehending subject and object RCs in Cantonese and Mandarin for each error type.

As in Kidd et al. (2015), only the head errors and reversal errors were analyzed, because, unlike the “other” errors, the processing strategies children use when making these two error types are readily interpretable. Since language is nested under the trilinguals (Cantonese, English, and Mandarin) but not the monolinguals, we compared the trilingual vs. monolingual Cantonese groups and the trilingual vs. monolingual Mandarin groups separately using the same analysis strategy. In addition, their head and reversal error responses were analyzed separately.

### Head Errors

#### Trilingual vs. monolingual Cantonese

We tested whether Extraction, Group (trilingual vs. monolingual Cantonese) and their interaction significantly contributed to head errors. We fitted a linear mixed effects model with the head error responses in trilingual Cantonese and monolingual Cantonese as the response. The fixed effects include Extraction, Group, and their interaction, and the significant random effects by subjects were also included in the model. By likelihood ratio tests, there was a significant effect of Extraction [ $\chi^2(1) = 36.98$ ,  $p < 0.001$ ] and a significant Group  $\times$  Extraction interaction [ $\chi^2(1) = 7.5$ ,  $p = 0.006$ ]. *Post-hoc* analyses that fit a linear regression model to analyze each extraction type separately showed that the group difference lay crucially in object but not subject RCs. Specifically, when comprehending Cantonese object RCs, the trilinguals made significantly more head errors than the monolinguals even though Cantonese is the first language for both groups [ $t_{(44)} = 2.44$ ,  $p = 0.02$ ]. When comprehending Cantonese subject RC sentences, the trilinguals and monolinguals did not exhibit a group difference [ $t_{(42)} = -1.35$ ,  $p = 0.18$ , *n.s.*].

### Trilingual vs. monolingual Mandarin

Similarly, we tested whether Extraction, Group (trilingual vs. monolingual Mandarin) and their interaction significantly contributed to head errors. We fitted a linear mixed effects model with the head error responses in trilingual Mandarin and monolingual Mandarin as the response. The fixed effects include Extraction, Group, and their interaction, and the significant random effects by subjects were also included in the model. By likelihood ratio tests, the only significant effect was Extraction [ $\chi^2(1) = 41.131, p < 0.001$ ], indicating that children made head errors significantly more often when comprehending Mandarin object RCs than Mandarin subject RCs. There was no significant effect of Group and it did not interact with Extraction, showing that the trilinguals had similar head noun error rate compared to their age-matched monolinguals when comprehending Mandarin RCs, a finding that is also consistent with comparing the two groups based on their correct responses.

### Reversal Errors

#### Trilingual vs. monolingual Cantonese

We tested whether Extraction, Group (trilingual vs. monolingual Cantonese) and their interaction significantly contributed to reversal errors. We used the same analysis strategy fitting a linear mixed effects model including random effects for subjects with the reversal error responses in trilingual Cantonese vs. monolingual Cantonese as the response, and Extraction, Group, and their interaction as fixed factors. By likelihood ratio tests, the only significant effect was Extraction [ $\chi^2(1) = 22.04, p < 0.001$ ], indicating that children made reversal errors significantly more often when comprehending Cantonese subject RCs (non-canonical VOS) than Cantonese object RCs (canonical SVO) in general. There was no significant effect of Group and it did not interact with Extraction, showing that the trilinguals and monolinguals were similar in terms of their tendency to make reversal errors when comprehending Cantonese RCs.

#### Trilingual vs. monolingual Mandarin

We tested whether Extraction, Group (trilingual vs. monolingual Mandarin) and their interaction significantly contributed to reversal errors. Likewise, we used the same analysis strategy to compare reversal error responses in trilingual Mandarin vs. monolingual Mandarin. The major results are similar to those comparing reversal errors in trilingual vs. monolingual Cantonese. The only significant effect was Extraction [Ext  $\chi^2(1) = 5.19, p = 0.02$ ], indicating that children made reversal errors significantly more often when comprehending Mandarin subject RCs (non-canonical VOS) than Mandarin object RCs (canonical SVO) in general. There was no significant effect of Group and it did not interact with Extraction, showing that the trilinguals and monolinguals were similar in terms of their tendency to make reversal errors when comprehending Mandarin RCs.

## Negative Transfer from L2 English to L1 Cantonese

To summarize, our error analyses revealed a crucial difference between the trilinguals and their age matched monolinguals when they comprehended Cantonese object RCs: the trilinguals made significantly more head errors than the monolinguals even though Cantonese is the first language for both groups. That is, the trilinguals were more likely to erroneously choose the subject of the RC as the head noun, choosing “mouse” instead of “chicken” as the head noun in (5) repeated as (14) below.

### Cantonese object classifier RC (head-final)

- (14) [RC 老鼠 錫\_\_\_\_] [<sub>headnoun</sub> 嗰 隻 公雞]  
 lou5syu2 sek3 go2 zek3 gung1gai1  
 mouse kiss that CL chicken  
 “The chicken that the mouse kisses”

We suggest that this group difference can be attributed to the trilinguals' knowledge of English, specifically these head errors in Cantonese could result from applying an English-based parsing strategy to the Cantonese object classifier RC stimuli. We will elaborate this argument further in the Discussion section.

## DISCUSSION

We have presented data involving the acquisition of two Chinese languages in a group of trilingual children who are also intensively exposed to English at school. The children from this study are acquiring the two Chinese languages under different exposure conditions, Cantonese as first language, Mandarin as their third language, under the heavy influence of English. We examined how these children's comprehension of subject and object RCs in the two Chinese languages is related to their knowledge of Cantonese, English and Mandarin. The results showed effects of both positive transfer and negative transfer across the three languages, showing bi-directional influence between the first and second/third languages. In particular, positive transfer from L1 Cantonese to L3 Mandarin allowed the trilingual children to comprehend Mandarin RCs above the level that would be expected based on their limited input. In contrast, negative transfer from L2 English to L1 Cantonese resulted in trilingual children having more difficulties in comprehending Cantonese object classifier RCs relative to their monolingual age peers.

Our hypotheses were that these trilinguals would experience facilitation in comprehending RCs in their L3 Mandarin; but would experience more difficulty in processing object classifier RCs in their L1 Cantonese relative to their monolingual peers. Our hypotheses are supported. In Mandarin, the trilinguals performed on a par with their monolingual age matched peers in comprehending complex sentences such as RCs, although Mandarin is their third and weaker language due to limited exposure. Recall these 5- to 6- year-old trilinguals scored lower than even the 3-year-old monolingual Mandarin children in terms of receptive vocabulary competence. In addition, their Cantonese RC performance and Mandarin RC performance

were strikingly similar (see **Figure 4**), leading us to argue that positive transfer from their first language Cantonese to their third language Mandarin is implicated. Our argument for positive forward transfer is further substantiated by showing that the trilinguals' L1 Cantonese RC performance uniquely positively predicts their L3 Mandarin RC performance, in particular for subject RCs. By contrast, their L1 Cantonese RC performance did not predict their L2 English RC performance, although the trilinguals exhibited a subject advantage across all the three languages.

In Cantonese, although having been extensively exposed to it from family and community since birth and their Cantonese receptive vocabulary scores are comparable to their age peers, the trilinguals performed significantly worse than the age-matched monolingual Cantonese group in comprehending Cantonese object classifier RCs because the trilinguals made significantly more head errors when parsing this construction. In these errors, the subject of the RC was mistakenly interpreted as the semantic head of a relative clause. These head errors could be a manifestation of negative influence from English, resulting from the trilinguals applying an English-based parsing strategy to the Cantonese object RC stimuli. Recall that Cantonese object RCs overlap with English head-initial subject RCs, in addition to overlapping with SVO transitive main clauses in all three languages. The trilinguals may have misparsed the Cantonese object RC stimuli using the English-based "head initial" analysis, erroneously taking the subject and the first noun (mouse), instead of the object and second noun chicken, as the semantic head of the relative clause, as shown in (15). A number of mechanisms may be implicated in the transfer of the English head-initial analysis. First, as described in the section Introduction, Cantonese object classifier RCs allow for an alternative internally headed analysis and this is a typological feature unique to Cantonese (see Chan et al., 2011 for more details). One possibility is thus that the head error arises from taking the subject to be the semantic head of an internally-headed relative clause.

- (15) [RC headnoun 老鼠 錫 緊 個 隻  
 lou5syu2 sek3 gan2 go2 zek3  
 mouse kiss PROG that CL  
 公雞] 喺 邊度 呀?  
 gung1gai1 hai2 bin1dou6 aa3  
 chicken is where SFP  
 (meaning: "where's the mouse that is kissing  
 the chicken?")

A second factor is that mis-parsing may be facilitated by the presence of the progressive aspect marker (PROG) *gan2* 緊 in the Cantonese stimuli which corresponds closely to the English suffix *-ing* [see (15) and **Table 4** for examples of test sentences]. Given this correspondence, it is possible that the trilinguals misparsed the Cantonese object RC stimuli similar to an English reduced subject RC [the mouse (that's) kissing the chicken], resulting in the head assignment error. Such English-based effects align with the fact that the trilingual and the monolingual groups differ crucially in terms of their exposure

to English. This interpretation of the error pattern predicts that children would make more head errors with Cantonese object RCs as their English dominance increased. We examined whether measures of language dominance or English proficiency would predict children's head noun errors in Cantonese object RCs in these trilinguals but found null results on this point. In a relevant study by Kidd et al. (2015, p. 447), however, a dominance effect was attested, as the study reported the main effect of dominance approaching significance in 20 simultaneous Cantonese-English bilinguals ( $p = 0.07$ ), suggesting that the children made fewer head errors as their Cantonese dominance increased. The difference in findings could be due to the fact that the bilingual children in Kidd et al. (2015) lived in an English-speaking environment (Canberra, Australia), and are thus likely to be more English dominant overall. We therefore concur with Kidd et al. (2015)'s suggestion that an important follow-up study would be to test a larger group of multilingual children with a wider array of dominance profiles.

One possibility as an alternative explanation is that these trilinguals were using an immature parsing strategy characteristic of younger monolingual Cantonese language learners. This alternative explanation is unlikely. First, although the trilingual and monolingual groups were not matched on their vocabulary scores, the trilinguals' performance in the L1 Cantonese vocabulary test was comparable to their age matched peers in the normed sample of the test, and the monolingual and trilingual groups in this study were age matched. Second, the trilinguals and monolinguals performed similarly on the Cantonese subject RCs, which for the monolinguals appear to be more difficult than object RCs. Third, even when we attempted to compare the trilinguals' performance profile in Cantonese with that of a younger group of monolingual Cantonese learners from another study reported in Chan et al. (2011), they are also distinctly different. Given that the trilingual group and the monolingual Cantonese group in the current study differ crucially in terms of their English exposure, and that their head errors were consistent with an English-based head-initial analysis, with structural overlaps between languages as a pre-condition, we therefore argue that the trilinguals' higher rate of head noun errors was more likely due to cross-linguistic influence from English.

The new findings complement and extend our previous works in a number of ways. They confirm our observation of cross-linguistic influence in Chinese-English bilingual's acquisition of RCs and extend this observation from simultaneous bilinguals to child second language acquisition. The presence of competing constructions makes SVO head-final object classifier RCs especially vulnerable in multilingual L1 Cantonese. Such vulnerability echoes the vulnerability reported in the L1 Cantonese of a group of simultaneous Cantonese-English bilingual children in Australia (Kidd et al., 2015). Our findings provide further evidence that head noun assignment in object classifier RCs is especially vulnerable to errors in multilingual Cantonese children under intensive exposure to English, even when they have been exposed to Cantonese as first language from birth. Our results therefore extend Kidd et al. (2015)'s observation of negative transfer from English to Cantonese

attested in a group of simultaneous bilinguals to another group of multilingual children who are also acquiring Cantonese under intensive exposure to English. Investigation of vulnerable linguistic domains in different multilingual child populations in relation to language exposure conditions and the language pair(s) involved can inform researchers about when and where cross-linguistic influence occurs in bilingual or multilingual development on one hand; and inform practitioners about when and where focused remediation may be considered on the other hand.

An additional empirical dimension offered by this study involves interactions between two Chinese languages (Cantonese and Mandarin) in multilingual child development. Specifically, in a trilingual Cantonese-English-Mandarin acquisition context, this study documents positive transfer from Cantonese to Mandarin between the trilinguals' first and third languages. Given that Mandarin is gaining prestige as a *lingua franca* among Chinese people in China, Hong Kong, Taiwan, Singapore and overseas communities, it is increasingly common for Chinese children to acquire one Chinese language as their home and first language, while also acquiring Mandarin simultaneously or successively from the community and/or school. These children develop some form of bilingualism involving two Chinese languages under different exposure conditions. The finding regarding positive transfer between the two Chinese languages also bears on the education of bilingual and trilingual children. Here we see the merit of children's L1 Cantonese benefiting the processing and acquisition of comparable structures in their L3 Mandarin despite limited exposure to Mandarin. We take this beneficial effect as a good reason to promote proficiency of Cantonese as heritage language for these trilingual children in their school and family education.

On the other hand, we also see interesting selective evidence of positive transfer from Cantonese to Mandarin in our trilingual children specific to one relative clause structure (subject relative) but not the other (object relative) that matches well with the similarities and differences between Cantonese and Mandarin relative clauses. Grammatical differences between Chinese languages and their implications for language acquisition have not received much attention so far. Investigating the acquisition of Chinese languages in bilingual/multilingual development requires recognizing that there are varieties of Chinese and considering the diversity in the specific properties of the target Chinese languages as an important linguistic factor in specifying where domains of facilitation or vulnerability may lie. For instance, we predict that positive transfer would not work for specific domains of grammar, such as acquiring the word order of double object datives in which Cantonese and Mandarin differ (Chan, 2010).

The current findings relate to theoretical perspectives in multilingual language acquisition and processing, especially with respect to cross-linguistic influences, in a number of ways. The positive transfer from L1 to the weaker L3 observed at such a young age is theoretically interesting from the perspective of psychotypology (the perception of linguistic distance). In the case of child language learners, age is associated with cognitive and metalinguistic development, and cognitive and metalinguistic

development could in turn be related to psychotypology: in general, one would expect that older children who have developed higher metalinguistic awareness may have a more accurate perception of linguistic distance. To the extent that psychotypology is possibly involved in the current group of 5- to 6- year-old trilinguals, it is impressive to observe young children having a perception of linguistic proximity between the two Chinese languages that could trigger forward positive transfer in processing certain similar syntactic structures. Taken together, then, in addition to the presence of structural parallels as a pre-condition for cross-linguistic transfer, positive transfer from L1 Cantonese to L3 Mandarin could be jointly driven by factors such as actual and perceived language distance (given the typological proximity between Cantonese and Mandarin) and language dominance (given that the trilinguals' Cantonese is more dominant than their Mandarin). Furthermore, recall the trilinguals' L1 Cantonese subject RC correct performance strongly and positively predicted their L3 Mandarin subject RC correct performance, and vice versa. This finding is also theoretically interesting because it constitutes suggestive evidence for shared syntactic representations between Cantonese and Mandarin in these young trilinguals and co-activation of their two typologically close languages during processing. The result is also consistent with psycholinguistic theories for bilinguals that posit shared syntactic representations between languages in instances of surface structure overlap (e.g., Meijer and Fox Tree, 2003; Hartsuiker and Pickering, 2008). To further test this hypothesis, a follow up study could be to test whether multilingual children acquiring Cantonese and Mandarin show any between-language priming effects between Cantonese and Mandarin (for subject RCs but not object RCs).

The finding regarding directionality of transfer from English to Cantonese is consistent with the prediction derived from Hulk and Müller's hypothesis. Recall that Cantonese object classifier RCs are potentially ambiguous between more than one analysis, and these Cantonese object classifier RCs overlap with subject RCs in English when the two languages are in contact in multilingual acquisition, while English RCs clearly allow only a head-initial analysis. According to Hulk and Müller's hypothesis, it would predict that Cantonese is the language being affected by cross-linguistic influence from English. Reverse transfer from L2 English to L1 Cantonese could therefore be triggered, especially when the children are under intensive exposure to English. The current finding further confirms the idea that if it is the structure in the first language that presents potential ambiguity of analyses, and the overlapping structure in the second language presents no ambiguity of analysis, reverse transfer from L2 to L1 is possible between two typologically divergent languages like Cantonese and English. In fact, such reverse transfer may be more likely to occur at an early age, during which the grammatical system of even the first language is under development for a multilingual child, making it more susceptible to cross-linguistic influence in vulnerable domains where structural ambiguity and competing analyses take time to resolve in the presence of structural overlaps. The current finding demonstrates that Hulk and Müller's hypothesis suffices to provide a unified theoretical perspective to jointly

consider cross-linguistic influence across bilingual and trilingual acquisition contexts.

A further remark about the effect of transfer. The structural overlap condition and Hulk and Müller's cross-linguistic influence hypothesis did not make specific predictions regarding whether the transfer is positive or negative. We view positive/negative transfer as an outcome rather than a process. The outcome depends on whether the overlapping analysis would lead to accurate usage/comprehension or errors/non-target forms in the target language. For subject RCs in Cantonese and Mandarin, the overlapping analysis leads to correct interpretation, hence positive transfer. For object classifier RCs in Cantonese, the overlapping analysis leads to incorrect interpretation, hence negative transfer. In the current case of negative transfer, we would like to further elaborate on the overlapping analysis, because the head-initial RC analysis preferred by the trilinguals is not permitted by the grammar of Cantonese. It is relevant to note that the Cantonese monolinguals also made this kind of head errors when comprehending object classifier RCs, although to a significantly lesser degree. This is not surprising in light of Hulk and Müller's cross-linguistic influence hypothesis and their notion of "vulnerable domain": cross-linguistic influence would occur in domains that are also known to be vulnerable and challenging for monolingual children. We further hypothesize that there is a coalition of "1st noun-as-agent" and "1st noun-as-head noun" processing preferences for young children in general (Bever, 1970; MacWhinney, 1977; Diessel and Tomasello, 2005). This general developmental tendency enables children to have good performance in comprehending subject RCs in languages with head-initial RCs like English and German (Diessel and Tomasello, 2005), but would give rise to developmental errors in head noun assignment for children acquiring head-final RCs, because "2nd N patient-as-head noun" conflicts with "1st N as agent and head noun" general processing preference. This hypothesis is further motivated by observing our published and unpublished data featuring languages like Cantonese, Mandarin and Dong with SVO head-final RCs that such kind of head noun assignment errors are not uncommon even among monolingual children (Yang and Chan, 2014; Kidd et al., 2015; Chan et al., 2017). Following this hypothesis, the overlapping analysis for our trilinguals in the current study would be the head-initial RC analysis, which is uniformly attested in English on one hand, and aligned with young children's general processing preferences on the other hand. This head-initial RC analysis therefore led to head noun assignment errors when the trilinguals comprehended Cantonese object classifier RCs, erroneously choosing the subject of the RC as the head noun. We can further view the mechanism of this negative transfer from a usage-based perspective (Tomasello, 2003; Lieven and Tomasello, 2008). The idea is that these developmental head noun assignment errors might be *more entrenched* in the trilingual children's Cantonese than those in the monolingual children. What is different between the trilingual and monolingual children's linguistic experience is that these trilingual children heard invariant head-initial RC forms in their additional and intensive English input. Apart from structural overlaps between object classifier RCs in Cantonese

and simple SVO transitive constructions in Cantonese and English, and subject RCs from English which are also SVO in order, there is also structural overlap between the invariant head-initial RC analysis from English and children's developmental tendency to choose the first mentioned noun phrase as the agent and the head noun of an SVO RC in Cantonese. As such, the tokens of head-initial RC forms in English that the trilingual children heard could have further entrenched these children's developmental tendency, making them increasingly accessible when it comes to syntactic choice in comprehension, leading to higher error frequency.

Limitations of the current study are also highlighted as follows. Trilingual language learners in the early years have three developing systems that can potentially influence each other. This study focused on documenting and accounting for the binary interactions between English and Cantonese (whereby English negatively influenced the parsing of a Cantonese grammatical construction), and the binary interactions between Cantonese and Mandarin (whereby Cantonese positively influenced Mandarin, with a possibility that Cantonese and Mandarin subject RCs have a shared representation in these trilinguals). Directionality of influence from English to Mandarin remains possible, but is difficult to test in the current case, mainly because if English were also influencing the non-dominant Mandarin, Mandarin is likely being jointly influenced by both Cantonese and English, and as such the joint influences cannot be teased apart. This study is therefore unable to investigate all the possible pathways of cross-linguistic influences between the three languages. Despite this, the current study points to an exciting new line of inquiry for future research. A fair amount of works on trilingualism have so far focused on how English as a lingua franca interacts with other languages in the European context (Cenoz and Jessner, 2000). As Mandarin becomes increasingly popular to acquire as a foreign language both for children and adults on a global scale, it will be extremely exciting to study how Mandarin is acquired as a L3 and how it interacts with other languages in a global context. A final remark about language dominance. While we are certain that Mandarin is the trilinguals' weakest language given their limited input (in contrast to Cantonese and English, which both featured prominently in these children's daily input), we are unsure about their relative dominance between Cantonese and English, as we have not used a comparable set of measures to systematically assess and compare these children's proficiency in Cantonese and English. Having only the receptive vocabulary scores from two different tests (CRVT and BPVS) does not allow us to make solid claims about the Cantonese-English dominance profiles of these trilinguals. The extent to which transfer from English to Cantonese is also driven by these trilinguals' dominance in English is unknown at the moment.

## CONCLUSION

This study is one of the very few studies that address cross-linguistic influences in young sequential trilingual children. We have identified a specific case of bidirectional influence between

the first and second/third languages in Cantonese-English-Mandarin trilingual children's comprehension of relative clauses. On the one hand, parallels and overlaps in both form and function provide a transparent basis for positive transfer from L1 Cantonese to L3 Mandarin, instantiating forward positive transfer from L1 to L3. On the other hand, intensive exposure to L2 English and structural overlaps in the languages cause multilingual children to experience more difficulty in processing object classifier RCs in their L1 Cantonese relative to their monolingual peers, instantiating backward negative transfer from L2 English to L1 Cantonese. These bi-directional cross-linguistic influences were attested within a single syntactic domain, demonstrating robust interactions between the linguistic systems of multilingual children. This study demonstrates how cross-linguistic interactions and exposure conditions could jointly influence acquisition outcomes: in this case, processing is facilitated by positive cross-linguistic influence despite limited exposure, and inhibited despite extensive exposure from birth due to negative cross-linguistic influence.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University. Ethics approval has been sought (reference number: HSEARS20150128006). Written

informed consent was also obtained from the parents of each participant.

## AUTHOR CONTRIBUTIONS

AC designed the experiment and interpreted the data in consultation with SM and VY. AC recruited the participants, supervised native-speaker experimenters in data collection, coded the data, supervised research assistants in data management such as reliability checks, and wrote a first draft of the paper. SC ran the statistical analyses. Subsequently all authors worked on refining and revising the text. All authors approved the final version.

## FUNDING

This research was partly supported by a research grant (G-YBF9; PI: Chan), awarded by The Hong Kong Polytechnic University. AC is a member of The Hong Kong Polytechnic University—Peking University Research Centre on Chinese Linguistics and its support is gratefully acknowledged.

## ACKNOWLEDGMENTS

We thank Ms. Ariel Chan and Mr. Francis Cho for help as research assistants, and all the children and schools who took part in this study.

## REFERENCES

- Ahukanna, J. G. W., Lund, N. J., and Gentile, J. R. (1981). Inter- and intra-lingual interference effects in learning a third language. *Modern Lang. J.* 65, 281–287. doi: 10.1111/j.1540-4781.1981.tb00985.x
- Anastassiou, F., and Andreou, G. (2014). "Multilingualism in Greece: a study of speech production of trilingual children," in *Major Trends in Theoretical and Applied Linguistics: Selected Papers from the 20th ISTAL*, Vol. 2, eds N. Lavidas, T. Alexiou, and A. Sougari (De Gruyter Open), 187–198. Available online at: <https://www.degruyter.com/viewbooktoc/product/422023>
- Bartelt, G. (1989). "The interaction of multilingual constraints," in *Interlingual Processes*, eds H. W. Dechert and M. Raupach (Tübingen: Gunter Narr.), 151–177.
- Bates, D., and Maechler, M. (2010). *Matrix: Sparse and Dense Matrix Classes and Methods. Rpackage Version 0.999375–39*. Available online at: <http://CRAN.R-project.org/package=Matrix>
- Bever, T. G. (1970). "The cognitive basis of linguistic structure," in *Cognition and the Development of Language*, ed J. R. Hayes (New York, NY: Wiley), 279–362.
- Cenoz, J., and Jessner, U. (eds.). (2000). *English in Europe: The Acquisition of a Third Language*. Clevedon: Multilingual Matters.
- Cenoz, J., Hufeisen, B., and Jessner, U. (eds.). (2001). *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters.
- Cenoz, J. (2001). "The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition," in *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*, eds J. Cenoz, B. Hufeisen, and U. Jessner (Clevedon: Multilingual Matters), 8–20.
- Chan, A. (2010). The Cantonese double object construction with *bei2* 'give' in bilingual children: the role of input. *Int. J. Bilingual.* 14, 65–85. doi: 10.1177/1367006909356653
- Chan, A., Lee, K., and Yip, V. (2014). "A new tool for assessing child mandarin receptive vocabulary," *Poster Presented at the 13th International Congress for the Study of Child Language* (Amsterdam).
- Chan, A., Matthews, S., and Yip, V. (2011). "The acquisition of relative clauses in Cantonese and Mandarin," in *The Acquisition of Relative Clauses. Processing, Typology and Function*, ed E. Kidd (Amsterdam, Philadelphia, PA: John Benjamins), 197–225.
- Chan, A., Yang, W. C., Chang, F., and Kidd, E. (2017). Four-year-old Cantonese-speaking children's online processing of relative clauses: a permutation analysis. *J. Child Lang.* doi: 10.1017/S0305000917000198. [Epub ahead of print].
- Cheung, P. S., Lee, K., and Lee, L. W. (1997). The development of the 'Cantonese Receptive Vocabulary Test for children aged 2–6 in Hong Kong. *Int. J. Lang. Commun. Disord.* 32, 127–138. doi: 10.3109/13682829709021465
- Cook, V. (ed.). (2003). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Diessel, H., and Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language* 81, 882–906. doi: 10.1353/lan.2005.0169
- Dryer, M. (2013). "Relationship between the order of object and verb and the order of relative clause and noun," in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: <http://wals.info/chapter/96> (Accessed September 18, 2017).
- Dunn, L. M., Dunn, L. M., Whetton, C., and Burley, J. (1997). *The British Picture Vocabulary Scale, 2nd Edn*. Windsor: NFER-Nelson.
- Dussias, P. E., and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingual. Lang. Cogn.* 10, 101–116. doi: 10.1017/S1366728906002847
- Edwards, M., and Dewaele, J. (2007). Trilingual conversations: a window into multicompetence. *Int. J. Bilingual.* 11, 221–242. doi: 10.1177/13670069070110020401

- Flynn, S., Foley, C., and Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *Int. J. Multilingual.* 1, 3–16. doi: 10.1080/14790710408668175
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gibson, E. (2000). "The dependency locality theory: a distance-based theory of linguistic complexity," in *Image, Language, Brain*, eds Y. Miyashita, A. Marantz, and W. O'Neil (Cambridge, MA: MIT Press), 95–126.
- Hao, M. L., Shu, H., Xing, A. L., and Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behav. Res. Methods* 40, 728–733. doi: 10.3758/BRM.40.3.728
- Hartsuiker, R. J., and Pickering, M. J. (2008). Language integration in bilingual sentence production. *Acta Psychol.* 128, 479–489. doi: 10.1016/j.actpsy.2007.08.005
- Helot, C. (1988). Bringing up children in English, French and Irish: two case studies. *Lang. Cult. Curr.* 1, 281–287. doi: 10.1080/07908318809525046
- Hoffmann, C. (1985). Language acquisition in two trilingual children. *J. Multilin. Multicult. Dev.* 6, 479–495. doi: 10.1080/01434632.1985.9994222
- Hoffmann, C., and Stavans, A. (2007). The evolution of trilingual code-switching from infancy to school age: the shaping of trilingual competence through dynamic language dominance. *Int. J. Bilingual.* 11, 55–72. doi: 10.1177/13670069070110010401
- Hsu, C. C. (2014). The role of age in Mandarin-speaking children's performance of relative clauses. *Concentric* 40, 29–54. doi: 10.6241/concentric.ling.40.2.02
- Hu, S., Gavarro, A., and Guasti, M. T. (2015a). Children's production of head-final relative clauses: the case of Mandarin. *Appl. Psycholinguist.* 37, 323–346. doi: 10.1017/S0142716414000587
- Hu, S., Gavarro, A., Vernice, M., and Guasti, M. T. (2015b). The acquisition of Chinese relative clauses: contrasting two theoretical approaches. *J. Child Lang.* 43, 1–21. doi: 10.1017/S0305000914000865
- Hulk, A., and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingual. Lang. Cogn.* 3, 227–244. doi: 10.1017/S1366728900000353
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- 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. Second Lang. Acquisit.* 29, 253–276. doi: 10.1017/S0272263107070131
- Kazzazi, K. (2011). Ich brauche mix-cough: cross-linguistic influence involving German, English and Farsi. *Int. J. Multiling.* 8, 63–79. doi: 10.1080/14790711003671879
- Keenan, E., and Comrie, B. (1977). Noun phrase accessibility and Universal Grammar. *Ling. Inq.* 8, 63–99.
- Kellerman, E. (1983). "Now you see it, now you don't," in *Language Transfer in Language Learning*, eds S. Gass and L. Selinker (Rowley, MA: Newbury House), 112–134.
- Kidd, E., Chan, A., and Chiu, J. (2015). Cross-linguistic influence in simultaneous Cantonese-English bilingual children's comprehension of relative clauses. *Bilingual. Lang. Cogn.* 18, 438–452. doi: 10.1017/S1366728914000649
- Kim, C. E., and O'Grady, W. (2015). Asymmetries in children's production of relative clauses: data from English and Korean. *J. Child Lang.* 43, 1038–1071. doi: 10.1017/S0305000915000422
- Lau, E. (2016a). Distinguishing ambiguity: Cantonese speakers using prosody to disambiguate surface identity in syntax. *J. Acoust. Soc. Am.* 140, 3396. doi: 10.1121/1.4970888
- Lau, E. (2016b). The role of resumptive pronouns in Cantonese relative clause acquisition. *First Language* 36, 355–382. doi: 10.1177/0142723716648840
- Li, G. (2006). Bilingual and trilingual practices in the home context: case studies of Chinese-Canadian children. *J. Early Childhood Literacy* 6, 355–381. doi: 10.1177/1468798406069797
- Lieven, E., and Tomasello, M. (2008). "Children's first language acquisition from a usage-based perspective," in *Handbook of Cognitive Linguistics and Second Language Acquisition*, eds P. Robinson and N. Ellis (Mahwah, NJ: Lawrence Erlbaum Associates), 168–196.
- Liu, H., Bates, E., and Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Appl. Psycholinguist.* 13, 451–484. doi: 10.1017/S0142716400005762
- MacWhinney, B. (1977). *Starting points.* *Language* 53, 152–168. doi: 10.2307/413059
- MacWhinney, B. (2005). "A unified model of language acquisition," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. Kroll and A. DeGroot (New York, NY: Oxford University Press), 49–67.
- MacWhinney, B. (2012). "The logic of a unified model," in *Handbook of Second Language Acquisition*, eds S. Gass and A. MacKey (New York, NY: Routledge), 211–227.
- Meijer, P. J. A., and Fox Tree, J. E. (2003). Building syntactic structures in speaking: a bilingual exploration. *Exp. Psychol.* 50, 184–195. doi: 10.1026/1617-3169.50.3.184
- Morett, L. M., and MacWhinney, B. (2013). Syntactic transfer in English-speaking Spanish learners. *Bilingual. Lang. Cogn.* 16, 132–151. doi: 10.1017/S1366728912000107
- Müller, N., and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingual. Lang. Cogn.* 4, 1–53. doi: 10.1017/S1366728901000116
- Oksa, E. (1977). "On becoming trilingual," in *Deutsch im Kontakt mit Anderen Sprachen*, ed C. Molony (Kronberg: Scriptor Verlag), 296–306.
- R Core Development Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>
- Reynell, J., and Huntley, M. (1987). *Reynell Developmental Language Scales—Cantonese (Hong Kong Version).* Windsor: NFER-Nelson.
- Ringbom, H. (1987). *The Role of the First Language in Foreign Language Learning.* Clevedon: Multilingual Matters.
- Rizzi, L. (1990). *Relativized Minimality.* Cambridge, MA: MIT Press.
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: determinants and mechanisms. *Ling. Approach. Bilingual.* 3, 3–25. doi: 10.1075/lab.3.1.01ser
- Stavans, A., and Muchnik, M. (2008). Language production in trilingual children: insights on code switching and code mixing. *Sociolinguist. Stud.* 1, 483–511. doi: 10.1558/sols.v1i3.483
- Stavans, A., and Swisher, V. (2006). Language switching as a window on trilingual acquisition. *Int. J. Multilin.* 3, 193–220. doi: 10.2167/ijm020.0
- Su, I. (2001). Transfer of sentence processing strategies: a comparison of L2 learners of Chinese and English. *Appl. Psycholinguist.* 22, 83–112. doi: 10.1017/S0142716401001059
- T'sou, B., Lee, T., Tung, P., Chan, A., Man, Y., and To, C. (2006). *Hong Kong Cantonese Oral Language Assessment Scale.* Hong Kong: Language Information Sciences Research Centre, City University of Hong Kong.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press.
- Yang, W.-C., and Chan, A. (2014). "When overlap with canonical word order is helping and hurting: evidence from bilingual children's comprehension of relative clauses in Dong and Mandarin," in *Poster presented at the 13th International Congress for the Study of Child Language* (Amsterdam).
- Yip, V., and Matthews, S. (2000). Syntactic transfer in a Cantonese-English Bilingual Child. *Bilingual. Lang. Cogn.* 3, 193–208. doi: 10.1017/S136672890000033X
- Yip, V., and Matthews, S. (2007a). Relative clauses in Cantonese-English bilingual children: typological challenges and processing motivations. *Stud. Second Lang. Acquisit.* 29, 277–300. doi: 10.1017/S0272263107070143
- Yip, V., and Matthews, S. (2007b). *The Bilingual Child: Early Development and Language Contact.* Cambridge: Cambridge University Press.

**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 © 2017 Chan, Chen, Matthews and Yip. 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.



# Home Language Will Not Take Care of Itself: Vocabulary Knowledge in Trilingual Children in the United Kingdom

Karolina Mieszkowska<sup>1\*</sup>, Magdalena Łuniewska<sup>1</sup>, Joanna Kołak<sup>1</sup>, Agnieszka Kacprzak<sup>1</sup>, Zofia Wodniecka<sup>2</sup> and Ewa Haman<sup>1</sup>

<sup>1</sup> Faculty of Psychology, University of Warsaw, Warsaw, Poland, <sup>2</sup> Institute of Psychology, Jagiellonian University, Kraków, Poland

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Monika S. Schmid,  
University of Essex, United Kingdom  
Francesca La Morgia,  
Trinity College, Dublin, Ireland

### \*Correspondence:

Karolina Mieszkowska  
karolina.mieszkowska@psych.uw.  
edu.pl

### Specialty section:

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

**Received:** 19 April 2017

**Accepted:** 25 July 2017

**Published:** 10 August 2017

### Citation:

Mieszkowska K, Łuniewska M,  
Kołak J, Kacprzak A, Wodniecka Z  
and Haman E (2017) Home  
Language Will Not Take Care of Itself:  
Vocabulary Knowledge in Trilingual  
Children in the United Kingdom.  
Front. Psychol. 8:1358.  
doi: 10.3389/fpsyg.2017.01358

Language input is crucial for language acquisition and especially for children's vocabulary size. Bilingual children receive reduced input in each of their languages, compared to monolinguals, and are reported to have smaller vocabularies, at least in one of their languages. Vocabulary acquisition in trilingual children has been largely understudied; only a few case studies have been published so far. Moreover, trilingual language acquisition in children has been rarely contrasted with language outcomes of bilingual and monolingual peers. We present a comparison of trilingual, bilingual, and monolingual children (total of 56 participants, aged 4;5–6;7, matched one-to-one for age, gender, and non-verbal IQ) in regard to their receptive and expressive vocabulary (measured by standardized tests), and relative frequency of input in each language (measured by parental report). The monolingual children were speakers of Polish or English, while the bilinguals and trilinguals were migrant children living in the United Kingdom, speaking English as a majority language and Polish as a home language. The trilinguals had another (third) language at home. For the majority language, English, no differences were found across the three groups, either in the receptive or productive vocabulary. The groups differed, however, in their performance in Polish, the home language. The trilinguals had lower receptive vocabulary than the monolinguals, and lower productive vocabulary compared to the monolinguals. The trilinguals showed similar lexical knowledge to the bilinguals. The bilinguals demonstrated lower scores than the monolinguals, but only in productive vocabulary. The data on reported language input show that input in English in bilingual and trilingual groups is similar, but the bilinguals outscore the trilinguals in relative frequency of Polish input. Overall, the results suggest that in the majority language, multilingual children may develop lexical skills similar to those of their monolingual peers. However, their minority language is weaker: the trilinguals scored lower than the Polish monolinguals on both receptive and expressive vocabulary tests, and the bilinguals showed reduced expressive knowledge but leveled out with the Polish monolinguals on receptive vocabulary. The results should encourage parents of migrant children to support home language(s), if the languages are to be retained in a longer perspective.

**Keywords:** trilingual language acquisition, trilingual children, multilingualism in migrant context, vocabulary acquisition, minority language, home language

## INTRODUCTION

The issue of how language input affects language acquisition in monolingual children has been a focus of broad scientific interest (e.g., Hart and Risley, 1995; Rowe, 2012; see Hoff, 2006 for review). Similarly, many studies have looked at how bilingual upbringing impacts the patterns of language input and how bilingual input influences language acquisition, especially in the area of vocabulary development (e.g., De Houwer, 2007; Gathercole and Thomas, 2009; Thordardottir, 2011; Hoff et al., 2012; Hoff and Core, 2013; Gollan et al., 2015; Unsworth, 2016).

The emerging field of investigating trilingual children's vocabulary acquisition has been largely dominated by case studies and has reported few comparisons with bilingual and monolingual performance. In the present paper we focus on trilingual children and explore their receptive and productive vocabulary in the community<sup>1</sup> language (English) and one of their home languages (Polish), in comparison with their bilingual and monolingual peers. We also investigate the properties of language input in trilingual children compared to bilinguals. We first briefly discuss what is known about the impact of language input on monolingual language acquisition and then present the available evidence on bilingual and trilingual language acquisition. As the issue of language development in trilinguals is still understudied, the rationale for the present analysis draws considerably on the evidence gathered from research on bilingual child development.

Research on monolingual language acquisition shows that quantity and quality of language input<sup>2</sup> in child's environment influence the pace of language development. In a ground-breaking study, Hart and Risley (1995) identified a group of monolingual children with diminished language input (caused indirectly by low family income and low parental education), who, at the age of 3, were estimated to hear 30 million fewer words than their peers from upscale families and had a significantly smaller vocabulary size. A follow-up study on the same children at the age of 9 revealed that the two groups grew further apart in their vocabulary knowledge and, accordingly, in their school performance, as measured by tests of listening, speaking, semantics, and syntax (Hart and Risley, 2003). The results of the studies by Hart and Risley show clearly that the amount of language input a child receives bears consequences for their language attainment and later school outcomes. Since then, researchers have further investigated the role of input in child language acquisition. Rowe (2012) discovered that the quantity of parental input alone was

insufficient in developing child's vocabulary at a preschool age, and identified that the diversity of parental vocabulary and use of decontextualized language (e.g., narratives) were the best predictors of pre-schoolers' vocabulary growth. Essentially, both quantity and quality of language input have been shown to influence the pace of child's language acquisition, including vocabulary development (Goodman et al., 2008; Unsworth, 2012, 2013).

Natural variation and diversity present in the language input of bilingual children may impact their vocabulary acquisition. In bilingual children, the quantity of input they receive is naturally divided between two languages, e.g., mother's vs. father's language, or L1 (i.e., home, heritage, or minority language) vs. L2 (i.e., community, or majority language). Thus, the nature of bilingual upbringing results in less input for each of the languages in comparison to the input received by monolingual peers (Pearson et al., 1993; Montrul, 2008, but cf. De Houwer, 2014). Reduced language input may be one of the reasons why bilingual children are repeatedly shown to score lower than monolinguals on vocabulary tasks in the majority language (e.g., Leseman, 2000; Oller et al., 2007; Bialystok et al., 2010; Bohnacker et al., 2016). Importantly, those vocabulary setbacks are found in different language pairs (e.g., Bialystok et al., 2010; Klassert et al., 2014; Bohnacker et al., 2016), across pre-school and school years (Bialystok et al., 2010), and – in the case of the majority language – are largely related to home-context vocabulary, rather than the school-context (Bialystok et al., 2010). A direct link between bilingual vocabulary development and language exposure was investigated by Thordardottir (2011) in a group of 5-year-old simultaneous French–English bilinguals in Canada<sup>3</sup>. Bilinguals' performance on receptive and expressive vocabulary was compared to that of their monolingual peers matched on age, socioeconomic status, and non-verbal intelligence, but differing in the amount of exposure they received in each language. A robust relationship was found between the amount of exposure to a language and children's performance in that language, although the relationship was observed to be different for the receptive and expressive vocabulary. Bilinguals exposed to both languages to the same extent scored comparably to monolingual children in the receptive vocabulary test, but they needed more input in a given language (and relatively less in the other one) to keep up with their monolingual peers in expressive vocabulary.

Access to many speakers of a given language seems to be another important factor contributing to language abilities in bilinguals. In a recent study by Gollan et al. (2015), the number of heritage language speakers that participants spoke to, correlated positively with their scores on a picture naming task (measured as the number of correct responses) in that language, and did not correlate negatively with their correctness in picture naming in English (community language). Importantly, the effect was independent of how frequently the participant used each language. Presumably, the greater the number of unique native speakers that a child interacts with on daily basis, the greater the

<sup>1</sup>In the present paper we use the terms "L1," "home language," "minority language," and "heritage language" interchangeably, and contrast those with "L2," "community language" and "majority language." In the Introduction, whenever we use a specific term, we follow the terminology chosen by the Authors of studies we report.

<sup>2</sup>In the present paper we use the terms "input" and "exposure" interchangeably when referring to the measurable interactions with the child in a given language. However, see Carroll (2017) for a call for a clearer definition of (and dissociation between) the terms.

<sup>3</sup>Though the bilingual context explored in the study by Thordardottir (2011) does not refer to bilingual migrants (the focus of this paper), the study investigates a direct link between language exposure to vocabulary skills in bilinguals and has thus been considered as relevant to the topic of the present paper.

variety of words used with a child, which may contribute to the child's vocabulary.

As demonstrated by the examples above, bilingual language development is a complex and dynamic process, influenced by, among others, the amount of input received in each language, and the number of native speakers of each language that the child has contact with. However, those factors vary in time and can change throughout the course of the child's development, resulting in shifts in language dominance. For instance, when bilingual upbringing is set in a migration context, the home language is usually the dominant one during the first years of the child's life. But when the child enters pre-school or school and the exposure to the community language increases, language dominance tends to shift toward the community language. In a questionnaire study aimed at determining factors that influence home language maintenance, De Houwer (2007) analyzed parental language use patterns from almost 2000 bilingual families, where at least one of the parents spoke a heritage language (different than the majority one). She asked how many of the children spoke the heritage language and found that nearly 25% of the children did not. De Houwer traced the origin to the parental language use patterns, showing most families spoke a mix of the heritage and community languages at home. Conversely, a model with the highest chances of successful home language maintenance was when at least one parent spoke only the heritage language at home. This is in line with the 20% threshold hypothesis (Pearson et al., 1997), which suggests that children who hear less than 20% of their input in a given language, are often reluctant to speak that language. According to Hoff et al. (2012), the 20% is an absolute minimum of input for a child to be able (and willing) to use a language. As established by Thordardottir (2011) in a previously mentioned study with 5 year olds in Canada, bilingual children achieved similar level of expressive vocabulary to that of their monolingual peers in either French or English, if they received 60% of their input in that language (French or English). Similar results are reported by Cattani et al. (2014) for children under the age of three exposed to English as a majority language.

Research on language development in trilingual children is an emerging field and features mostly case studies. Kimbrough Oller (2010), who analyzed all-day recordings from a toddler trilingual with German, English and Spanish, showed that directedness of input in the three languages was strongly predictive of the number of words that the child used in each language. Consequently, the child produced more words in the language that was spoken to her directly, compared to the language heard by the child, but not addressed to her. Hoffmann (1985, 2001) spent 7 years observing two early trilingual children, both of whom acquired Spanish and German from their parents, and English, their third language, from the community, school, and peers. The study showed that the children developed "sufficient competence in all three languages to fulfill their communication needs as they were at the time" (Hoffmann, 2001, p. 3). Montanari (2009) found that a Tagalog-Spanish-English trilingual child was able to select the appropriate language according to the interlocutors' linguistic repertoire before the age of two and that the occasional instances of inappropriate language use

were mostly due to vocabulary gaps. Observing the same child (Montanari, 2010), she found that the child's cumulative vocabulary growth from 1;4 to 2;0 was fairly comparable to that of bilinguals' and monolinguals' tested by Pearson et al. (1993). The conclusions from the case-studies of trilinguals are in-line with research on bilingual development, but there is still a need for more extended investigations on larger samples.

Our goal was to examine the vocabulary knowledge in migrant children (lower primary school) who have frequent contact with three languages, and to map the outcomes of vocabulary tests onto the patterns of language use reported by children's parents. The specific aims were the following:

- (1) To investigate the vocabulary knowledge in migrant trilingual children and compare it with the knowledge of their bilingual and monolingual peers.
- (2) To explore the relative frequency of input in each language in trilingual families and contrast these patterns with those in bilingual families.
- (3) To link language input to vocabulary knowledge in multilingual children.

We explored the actual performance on the receptive and expressive vocabulary tests in trilingual children, and compared those with the lexical performance of bilingual and monolingual peers. We then viewed those results in the light of relative frequency of input in each language in trilingual children and compared it with the input reported in bilingual peers.

## MATERIALS AND METHODS

### Participants

We analyzed data gathered in a larger project on cognitive and language development of Polish bilingual children (related to COST Action IS0804<sup>4</sup>). The database collected in the Bi-SLI-PL project consists of data from 173 bilingual children living in the United Kingdom who had at least one Polish parent, 311 Polish monolingual children, and 30 English monolingual children. A written parental consent was obtained for all the children participating in the study. In addition to the vocabulary testing, participants completed a large battery of tools measuring grammar knowledge, phonological processing and storytelling, however, the results of these tests are beyond the scope of this paper.

For the current analyses, we used data from 56 children, trilingual, bilingual, and monolingual. We first selected all children who had been exposed to more than two languages ( $n = 14$ ). These children, i.e., the trilingual group living in the United Kingdom, were born to families with one Polish parent and one parent of other nationality, so they were exposed to two home languages from birth: Polish and another language (Albanian/Arabic/Bengali/French/Italian/Macedonian/Russian/Ukrainian), and to the majority language,

<sup>4</sup>www.bi-sli.org

English (age of onset:  $M = 8$  months,  $SD = 14$  months, range: 0–36 months). The selected group of participants was matched (in a one-to-one pairwise fashion) with the peer groups of: (1) Polish–English bilinguals living in the United Kingdom ( $n = 14$ ); (2) Polish monolinguals living in Poland ( $n = 14$ ); (3) English monolinguals living in the United Kingdom ( $n = 14$ ). The pairwise matching was based on the chronological age, gender, and the non-verbal intelligence score. We also compared the children's socio-economic status (measured in the years of maternal education). A Kruskal–Wallis test showed no statistically significant difference in SES between the four groups,  $H(3) = 5.7$ ,  $p = 0.125$  (see **Table 1** for details).

## Procedure

The data analyzed here were gathered in the Bi-SLI-PL project (related to COST Action IS0804). The project used a number of measures of linguistic and cognitive development (see Haman et al., Unpublished). The present analysis focuses on vocabulary measures (receptive and expressive vocabulary size) and the parental reports of the child's input in each of their languages.

## Expressive and Receptive Vocabulary Tests

We used standardized picture-naming and word-recognition tests in Polish and English in the case of bilinguals and trilinguals, or in one of those languages in the case of monolinguals. For English, we applied the Expressive Vocabulary Test (Williams, 2007) to assess the children's expressive word knowledge, and for Polish we used Zadanie Nazywania Obrazków (Haman and Smoczyńska, 2010, Unpublished). In both tests of expressive vocabulary we asked the children to name pictures illustrating objects (for nouns as target words), their features (adjectives), or some activities (verbs). Receptive word knowledge was assessed with the British Picture Vocabulary Scale (BPVS-3; Dunn et al., 2009) in English, and Obrazkowy Test Słownikowy – Rozumienie (OTSR; Haman et al., 2012) in Polish. In both tests of receptive vocabulary children were asked to choose one picture depicting the target word out of four colorful pictures presented on each board. The raw scores from the tests were transformed into standard scores (z-scores). The mean score and the standard deviations were calculated on the monolingual populations (monolingual Polish for the Polish vocabulary tests, and monolingual English for the English vocabulary tests). Using standard scores allowed us to establish how far from the monolingual mean were the scores of the bilingual and trilingual groups.

## Parental Reports of Input in Each of the Child's Languages

We used a Polish version of the Questionnaire for Parents of Bilingual Children<sup>5</sup> [(PABIQ – Tuller, 2015; Polish adaptation by Kuś et al. 2012, Unpublished)] to extract the information about the number of speakers and the frequency of bilingual and trilingual children's input in the home and majority languages<sup>6</sup>. Specifically, we asked parents to estimate on a five point Likert scale how often (and with whom) their child was addressed in each language in specific communicative situations in two types of settings: at home and outside of home.

The communicative situations at home (henceforth referred to as at-home input) included two factors with different weights: we asked the parents to estimate how often each language was used toward the child by each of the parents and the siblings [from 0 = “never,” 2 = “rarely,” 4 = “sometimes,” 6 = “most of the time,” 8 = “always,” maximum score: three sources (mother, father, siblings) \* 8 “always” = 24 points]. We also asked them to specify how often each language was used toward the child by the grandparents and the possible care-takers (e.g., babysitter), and was used in the parent-to-parent interaction (i.e., language not directed toward the child but which can still be overheard by the child) [from 0 = “never,” 1 = “rarely,” 2 = “sometimes,” 3 = “most of the time,” 4 = “always,” maximum score: four sources (grandparents, babysitter, mother speaking to father, father speaking to mother) \* 4 “always” = 16 points]. Thus, the input from the parents and siblings was weighted more than the input from other adults close to the family. The maximum total score on the index of at-home input was 40 points (24 + 16) for each language. The higher the number of the speakers in a particular language, the higher the total score of input in that language (accordingly, the total score was proportionately lower if child did not have contact with their grandparents and/or did not have siblings). To allow an approximate assessment of the

<sup>5</sup>The same questionnaire was used in Haman et al. (Unpublished), to calculate an index of cumulative exposure to L1 and L2 in bilingual children. However, for the present analysis we calculated a different index related to the frequency of input in each of the multilingual child's languages, and the number of speakers in the child's linguistic environment.

<sup>6</sup>Interpretation of the questionnaire data need to be treated with caution: the parental estimations of frequency of input, as any self-reported measures, are by nature subjective and intuitive. This point is taken up in more detail in Study Limitations. However, we would like to note here that the same indices of language input, or parallel indices based on the same set of questions, have been used by researchers before in studies on different languages (e.g., Blom and Bosma, 2016; Bohnacker et al., 2016; dos Santos and Ferré, 2016; Fleckstein et al., 2016; Rinker et al., 2017).

**TABLE 1 |** Characteristics of the participants: gender, age (in months), non-verbal intelligence score and maternal education (in years) across trilinguals, bilinguals, Polish monolinguals, and English monolinguals.

|                                                    | Trilinguals<br>( $n = 14$ ) | PL-EN bilinguals<br>( $n = 14$ ) | PL monolinguals<br>( $n = 14$ ) | EN monolinguals<br>( $n = 14$ ) | Between-groups<br>comparison     |
|----------------------------------------------------|-----------------------------|----------------------------------|---------------------------------|---------------------------------|----------------------------------|
| Gender                                             | 6 m + 8 f                   | 6 m + 8 f                        | 6 m + 8 f                       | 5 m + 9 f                       | $\chi^2(3) = 0.22$ , $p = 0.975$ |
| Age (months) $M \pm SD$                            | 66 $\pm$ 9                  | 66 $\pm$ 9                       | 67 $\pm$ 9                      | 66 $\pm$ 9                      | $H(3) = 0.03$ , $p = 0.999$      |
| Raven (raw score) $M \pm SD$                       | 23 $\pm$ 5                  | 23 $\pm$ 4                       | 23 $\pm$ 5                      | 21 $\pm$ 4                      | $H(3) = 2.18$ , $p = 0.537$      |
| Maternal education (years of schooling) $M \pm SD$ | 15 $\pm$ 3                  | 17 $\pm$ 3                       | 16 $\pm$ 3                      | 17 $\pm$ 2                      | $H(3) = 5.7$ , $p = 0.125$       |

relative contribution of each language into a child's language input, the total score in each language was transformed into a percentage value. For instance, to get a percentage value for Polish, we divided the total score for Polish by the sum of the total scores for all child's languages, and multiplied it by 100<sup>7</sup>.

The communicative situations outside of home (henceforth referred to as outside-of-home input) included a number of factors with different weights: the number of hours spent at school divided by 3 (maximum score: 36 h/3 = 12 points), participation in after-school activities in each language (from 0 = "never," 1 = "once a week," 2 = "everyday," maximum score: 2 points), frequency of book reading, storytelling, rhymes/singing, computer games, TV/movies watching in each language (from 0 = "never," 1 = "once a week," 2 = "everyday," maximum score: five activities \* 2 "everyday" = 10 points). Parents also estimated how often each language was used toward the child by their peers (from 0 = "never," 2 = "rarely," 4 = "sometimes," 6 = "most of the time," 8 = "always," maximum score: 8 points), and by family guests and/or relatives not living in the house [from 0 = "never," 1 = "rarely," 2 = "sometimes," 3 = "most of the time," 4 = "always," maximum score: two sources (family guests, relatives) \* 4 "always" = 8 points]. The maximum total score for the frequency of outside-of-home input was 40 points for each language. The higher the number of activities/additional speakers in a particular language, the higher the total score of input in that language (accordingly, the total score was proportionately lower if child did not attend any extracurricular/listed activities, or the family did not have any regular visitors). The total scores in each language were transformed into percentage values (i.e., to get a percentage value for Polish, we divided the total score for Polish by the sum of the total scores for all child's languages, and multiplied it by 100).

### Testing Procedure

The children were tested by a native or near-native speaker of the language (Polish or English) in a quiet room: the monolingual children in their preschools, the bilingual and trilingual children in their day-cares, schools or in their homes in the United Kingdom. The bilingual and trilingual children were tested by different experimenters, and on different days in each of their respective languages. They were first tested in their dominant language (either Polish or English, as reported by the parents), and then in the other language. There was a maximum of a 6-week break between the two language testing sessions.

### Statistical Analysis

Given the small size of each sample ( $n = 14$ ), we employed non-parametric tests of group differences to tackle potential violation of normality assumption in ANOVA. We performed a series of Wilcoxon–Pratt Signed-Rank Test to compare amount of contact with home and majority languages between bilingual and trilingual children. We used Kruskal–Wallis

tests to contrast the receptive and expressive vocabulary knowledge of trilingual, bilingual, and monolingual children. Whenever the Kruskal–Wallis tests revealed significant differences between the groups, we used Nemenyi test as *post hoc*.

## RESULTS

### Vocabulary of Trilinguals

The main aim of the current analysis was to examine the vocabulary knowledge in trilingual children in comparison with the bilingual and monolingual groups.

#### Vocabulary in English

The trilinguals' raw scores on receptive and productive vocabulary tests in English were compared to those of the bilinguals and English monolinguals. The descriptive results are presented in **Figure 1**.

We used the Kruskal–Wallis test to compare the English receptive vocabulary scores across the three groups. We found no significant effect of group,  $H(2) = 4.81$ ,  $p = 0.09$ .

The Kruskal–Wallis test used to compare English productive vocabulary scores across the three groups showed a marginally significant effect of group,  $H(2) = 6$ ,  $p = 0.049$ . However, a Nemenyi *post hoc* revealed no significant difference between the groups (trilinguals vs. bilinguals:  $p = 0.956$ , trilinguals vs. English monolinguals:  $p = 0.119$ , bilinguals vs. English monolinguals:  $p = 0.062$ ).

#### Vocabulary in Polish

The trilinguals' raw scores on the receptive and productive vocabulary tests in Polish were compared to those of the bilinguals and Polish monolinguals. The descriptive results are presented in **Figure 2**.

The Kruskal–Wallis test was used to compare the Polish receptive vocabulary scores across the three groups, and showed a significant effect of group,  $H(2) = 8.23$ ,  $p = 0.016$ . A *post hoc* analysis revealed that the Polish monolinguals scored significantly higher on the receptive test in comparison with the trilinguals ( $p = 0.012$ ). We found no significant differences between the receptive Polish vocabulary scores of the Polish monolinguals and bilinguals ( $p = 0.279$ ). Furthermore, we found no statistically significant difference between the scores of the bilinguals and trilinguals ( $p = 0.373$ ).

Again, we used the Kruskal–Wallis test to compare the Polish productive vocabulary scores across the three groups. We found a significant effect of group,  $H(2) = 21.89$ ,  $p = 0.001$ . A *post hoc* analysis revealed that Polish monolinguals scored significantly higher on productive vocabulary in comparison with trilinguals ( $p = 0.001$ ), and bilinguals ( $p = 0.013$ ). Though the trilinguals' average score was numerically lower than that of the bilinguals and Polish monolinguals, we found no statistically significant difference between the scores of the trilinguals and bilinguals ( $p = 0.166$ ) on the productive vocabulary size in Polish.

<sup>7</sup>It is important to note, that the indices described above do not indicate the absolute amount of input received in each of the child's languages, but rather the relative frequency of input received in each language, which is, among others, influenced by the number of speakers in child's linguistic environment, and regular activities performed in each language.

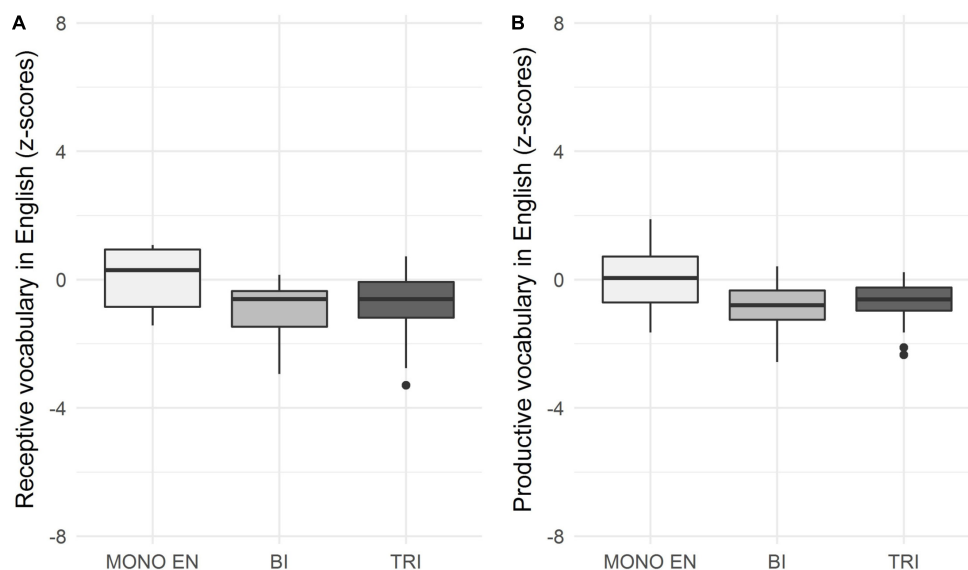


**FIGURE 1 |** Receptive (A) and productive (B) vocabulary in English (z-scores) across three groups: monolingual English, bilingual, and trilingual children.

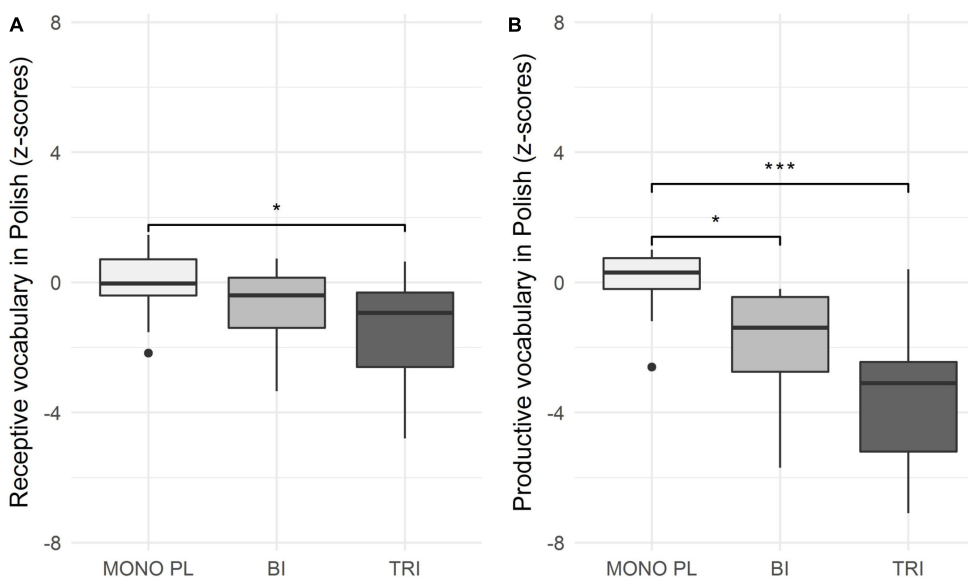


**FIGURE 2 |** Receptive (A) and productive (B) vocabulary in Polish (z-scores) across three groups: monolingual Polish, bilingual, and trilingual children. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .

## Language Use Patterns in Bilingual and Trilingual Families

In order to examine the vocabulary results in view of the language environment of our participants, we compared the frequencies of input in each language in the bilingual and trilingual groups. We focused our comparison on those two groups because we were interested specifically in the language use patterns in the bilingual and trilingual families. In the case of Polish and English monolingual children, we assumed their input was wholly in their native language. The descriptive results from the bilingual

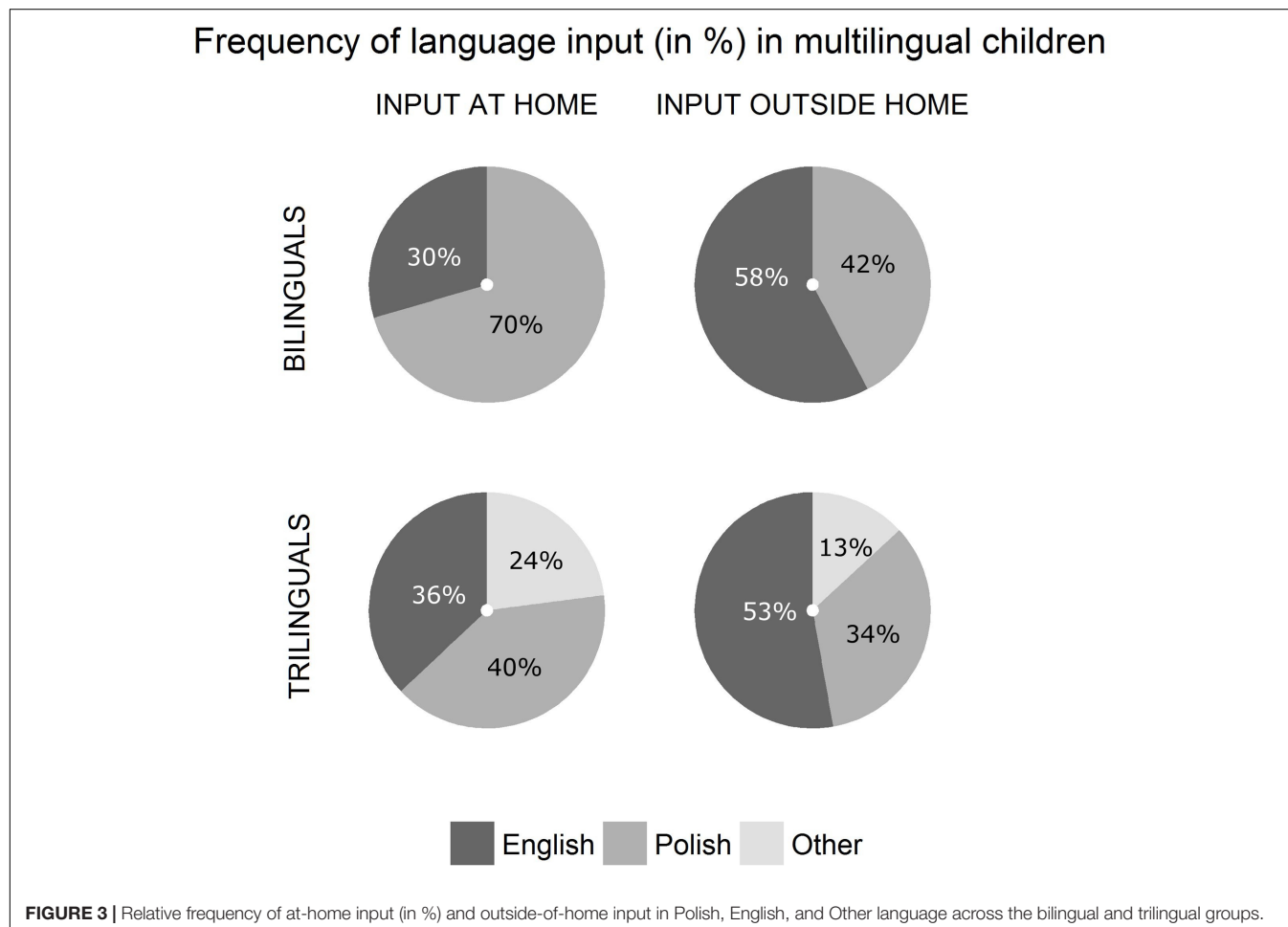
and trilingual groups are given in **Table 2** and presented in **Figure 3**.

### At-home Input

While the bilinguals' at-home input was predominantly Polish, the frequency of the trilinguals' at-home input was more equally distributed between the three languages: Polish, English and Other (see **Table 2** and **Figure 1**). Wilcoxon–Pratt Signed-Rank Test showed that the trilinguals and bilinguals differed significantly in the relative frequency of input in Polish ( $W = 15$ ,

**TABLE 2 |** The frequency of at-home input and outside-of-home input (in %) in each language across bilingual and trilingual groups.

|                           | Bilinguals                          |                                    | Trilinguals                         |                                    |                                   |
|---------------------------|-------------------------------------|------------------------------------|-------------------------------------|------------------------------------|-----------------------------------|
|                           | English<br>( <i>M</i> ± <i>SD</i> ) | Polish<br>( <i>M</i> ± <i>SD</i> ) | English<br>( <i>M</i> ± <i>SD</i> ) | Polish<br>( <i>M</i> ± <i>SD</i> ) | Other<br>( <i>M</i> ± <i>SD</i> ) |
| At-home input (%)         | 30 ± 19<br>Range: 0–59              | 70 ± 19<br>Range: 41–100           | 36 ± 18<br>Range: 0–61              | 40 ± 13<br>Range: 16–63            | 24 ± 11<br>Range: 11–48           |
| Outside-of-home input (%) | 58 ± 8<br>Range: 43–74              | 42 ± 8<br>Range: 26–57             | 53 ± 7<br>Range: 41–71              | 34 ± 7<br>Range: 17–44             | 13 ± 6<br>Range: 4–21             |



$Z = -2.75$ ,  $p = 0.003$ ), with the trilinguals hearing Polish less frequently, relative to bilinguals. However, there was no difference between the groups on the frequency of input received in English ( $W = 102.5$ ,  $Z = 0.7$ ,  $p = 0.519$ ).

#### Outside-of-Home Input

The two groups heard English spoken in the outside-of-home context equally frequently ( $W = 49$ ,  $Z = -1.49$ ,  $p = 0.151$ ). Also, outside of home, the two groups heard English more frequently than any other language. However, the Wilcoxon–Pratt Signed-Rank Test showed that the two groups differed significantly in the frequency of the outside-of-home input in Polish ( $W = 38$ ,  $Z = -2.43$ ,  $p = 0.012$ ),

with the trilinguals hearing Polish less frequently than the bilinguals.

Overall, the results on language use patterns in bi- and trilingual homes reveal that while the two groups heard English equally frequently, they differed significantly in the frequency of the input in Polish, with the trilinguals having less frequent contact with Polish than the bilinguals.

#### Linking Vocabulary Scores and Frequency of Input

Finally, we investigated the relationship between the vocabulary scores and the relative frequency of the input received in English and Polish. For this purpose, we used a combined

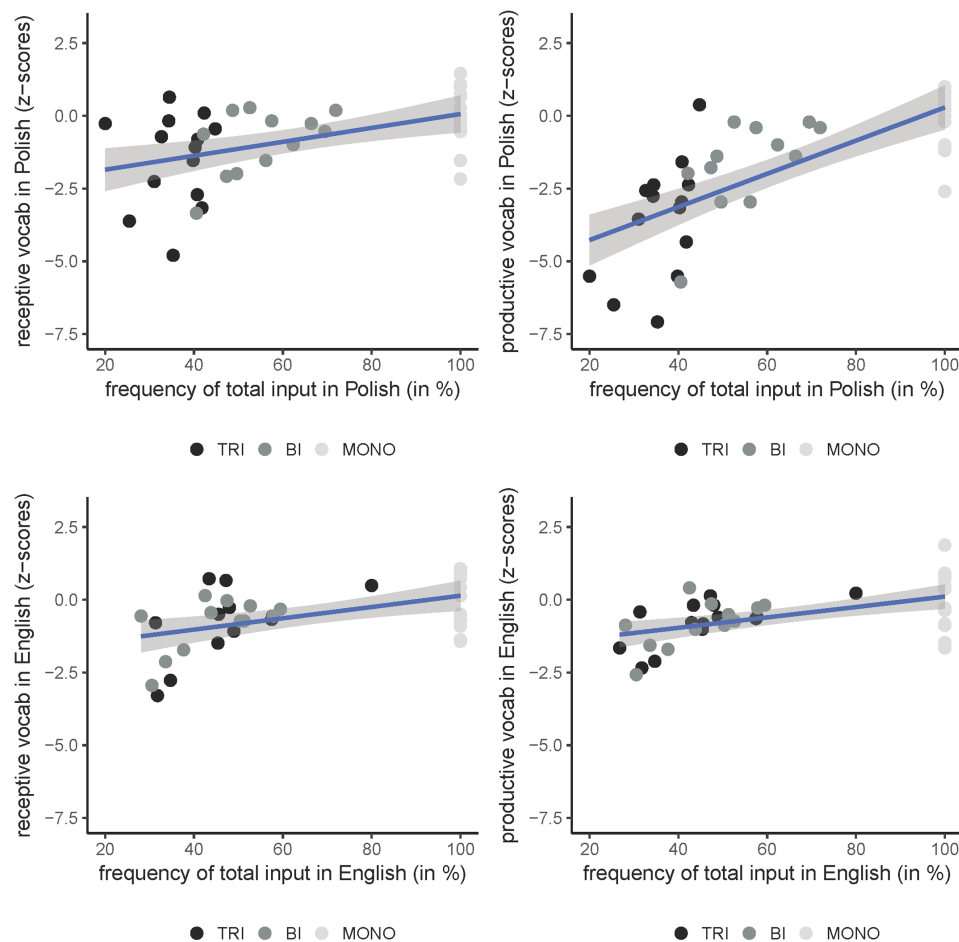


**FIGURE 4 |** The distribution of vocabulary scores (receptive and productive) of individual children from the three groups in relation to the relative frequency of total input (at home and outside of home) received in English and Polish. The lines represent lines of best fit with the 95% confidence level interval.

index of language input which was a sum of the input at-home and outside-of-home. A series of Spearman's rank correlations assessed the relationship between the vocabulary scores in English and Polish and the relative frequency of the input received in the two languages. The correlations were done on data from all the subjects, with no differentiation between the trilingual, bilingual, and monolingual groups. **Figure 4** presents the correlations, separately for English and Polish and for the receptive and productive vocabulary scores. In English, the correlation between the relative frequency of total input received in English and the vocabulary scores was  $r_s = 0.539$ ,  $p < 0.001$  for the productive vocabulary, and  $r_s = 0.451$ ,  $p < 0.01$  for the receptive vocabulary. For Polish, the correlation between the relative frequency of the input in Polish and the vocabulary score was  $r_s = 0.821$ ,  $p < 0.001$  in the productive test, and  $r_s = 0.503$ ,  $p < 0.001$  in the receptive test.

Overall, the results show that the relative frequency of the input received in each language was positively and strongly correlated with the vocabulary scores in this language. In both languages, the correlations were stronger in the domain of the productive vocabulary.

## DISCUSSION

The goal of the paper was to compare lexical knowledge of trilinguals with their bilingual and monolingual peers and to relate the vocabulary outcomes to the daily patterns of language use reported by the parents. To this aim, we compared the results in their expressive and receptive vocabulary tests with those of carefully matched bilingual and monolingual children. We also analyzed their language outcomes in the light of the relative frequency of input in each of the trilingual children's languages, and compared them with the input received by their bilingual counterparts.

First, we examined the vocabulary knowledge of trilinguals, Polish-English bilinguals, and Polish and English monolinguals. We compared their receptive and expressive vocabulary scores for Polish (the home language) and English (the community language). For English, the results revealed no significant differences between the bilinguals' and trilinguals' vocabulary size on either the receptive, or productive vocabulary tests. Moreover, the two groups did not differ from the English monolinguals in their receptive and productive vocabularies.

For Polish, however, the results paint a much more complex picture. The trilinguals and bilinguals showed similar vocabulary scores in both receptive and productive tests in Polish. When we compared the vocabulary knowledge of the two groups with that of Polish monolinguals, differences occurred. With respect to the receptive vocabulary in Polish, we found that while the bilinguals did not differ from the Polish monolinguals, the trilinguals had significantly smaller receptive lexicons. In terms of the productive vocabulary in Polish, both the bilinguals and trilinguals scored significantly lower than the Polish monolinguals. Additionally, we investigated the relationship between the children's vocabulary scores and the relative frequency of the input received in English and Polish. We found positive strong correlations between the relative frequency of the input received in each language and the vocabulary scores in that language. In both languages, the correlations were stronger in the domain of the productive vocabulary.

Overall, we found two main characteristics of trilingual development in the immigrant context. First, we established that in terms of the majority language (English), the trilingual and bilingual children in our sample showed vocabulary knowledge similar to that of their monolingual peers. This was demonstrated by the lack of significant differences between the three groups in both receptive and productive English vocabulary tests. The current data provide evidence in support of the claim expressed by Grosjean that in a migrant context, the majority language is likely to take care of itself, mostly due to the large exposure to the language in the daycare or school and from the peers (Grosjean, 2010, p. 209). This we have found to be equally true for both the bilingual and trilingual children in our sample.

Secondly, we found that the bilinguals and trilinguals showed significantly lower vocabulary knowledge in their home language Polish, as compared to non-migrant Polish monolingual peers matched on age, gender, and non-verbal IQ. A possible explanation of this finding may lie in the patterns of language use in the multilingual homes. Since the quantity of language input in the child's environment has been long established as crucial for the pace of language development (Goodman et al., 2008; Rowe, 2012; Hoff, 2013; Gollan et al., 2015), receiving less input in the home language may cause setbacks in developing comprehension and production in that language. To explore the potential impact of language input on the vocabulary knowledge in our sample, we examined the relative frequency of the input in each language in the bilingual and trilingual groups. To this end, the parents of bilingual and trilingual children were asked to specify how often the child is addressed in each of the languages at home (i.e., among family members) and outside of home (i.e., by peers, at school, during after-school activities). The analysis of the questionnaire data revealed that the trilinguals and bilinguals heard English (the majority language) equally frequently at home. Additionally, both groups heard English most frequently outside of home. However, the bilinguals and trilinguals differed significantly in the reported frequency of input in Polish (the home language). The trilinguals in our sample heard Polish less frequently both at home and outside of home, as compared to the bilinguals.

The data we gathered do not reflect the absolute amount of input the children received in each of their languages, rather the relative frequency of input received in each language. Nevertheless, we have found that the relative frequency of the input in the home language (Polish), naturally reduced in the bilingual context, was even further limited in the trilingual home, where communication was divided between the two parental languages. Against this background, we have seen a worse performance of the trilingual and the bilingual groups on the Polish vocabulary tests, relative to their Polish monolingual peers.

The present analysis has practical implications for the parents and caretakers of bilingual and multilingual children. The results indicate that the majority language may develop equally well to that of monolinguals, but it is the home language(s) that require(s) more considerate attention. It seems that when the bilingual and the trilingual children enter preschool or school, the exposure to the community language increases, shifting language dominance toward that language. This may eventually lead to developing language abilities predominantly in the community language at the expense of the home language, especially if the home language enjoys lower social prestige than the majority language (Gathercole and Thomas, 2009). To maintain their home languages children need rich and varied home language input. Since the nature of bilingual and trilingual upbringing results in getting less input for each of the languages in comparison to the input received by monolinguals, it is important to maintain the quality of the child directed input. Other studies have demonstrated that of crucial importance to the child's developing lexicon is not only the mere quantity of input, but also diversity of vocabulary used (e.g., Rowe, 2012), utterance length (Hoff, 2003), and the number of speakers (Gollan et al., 2015). Some researchers, e.g., De Houwer (2007) suggest multilingual families should have at least one parent speaking only the home language, with no code-mixing.

Thus, we would like to call attention to the quality of input a multilingual child receives and we believe it is important to encourage parents and practitioners to invest in all sorts of child-friendly activities (play groups, reading clubs, etc.) as to provide linguistically rich and varied input of the home language(s) in out-of-home context and more opportunities to use the language(s).

It is crucial to stress that the amount and quality of language input a bilingual or trilingual child receives bears an impact on their attainment of the home language. This might turn out particularly important in view of return migration to the parents' home country, where children often experience educational difficulties, mostly due to the fact that their home language is relatively weaker than their former majority language (Grzymała-Moszczyńska et al., 2015).

## Study Limitations

The presented analysis is not without limitations. The first one is a small size of the compared groups (each group consisted of only 14 children). However, it needs to be stressed that the compared groups were carefully selected and matched. First, to ensure the groups' homogeneity on potential confounding distracting

factors, we employed pair-wise matching to the trilingual group on multiple variables (chronological age, gender, and non-verbal intelligence score). Secondly, the groups did not differ in either age, gender, non-verbal IQ score, or socio-economic status (measured in years of maternal education). Moreover, since our data did not follow normal distribution, we used non-parametric tests for the analyses, the statistical power of which is weaker, i.e., they are less likely to find statistical differences.

Another constraint of the study is the interpretation of the questionnaire data concerning the frequency of language input: the estimation provided by the parents are by definition non-objective and intuitive (e.g., when filling in the questionnaire, parents choose whether a particular language is used toward the child “sometimes” or “most of the time” based on their own interpretations of the scale). Moreover, our index does not account for the variance in the amount of parent–child contact between mothers and fathers (it is possible that the mothers had relatively more contact with the children than the fathers, e.g., the mother’s frequent use of Polish may not equal the father’s frequent use of another language). Nevertheless, the same indices of language input were repeatedly used before (e.g., Blom and Bosma, 2016; Bohnacker et al., 2016; dos Santos and Ferré, 2016; Fleckstein et al., 2016; Rinker et al., 2017) as valid measures of the relative frequency of input in the child’s languages.

Finally, the participants were tested only at one study point. Therefore we were not able to observe the changes in their access to input in the languages over time and the potential changes in their lexical knowledge. Such an analysis would be most informative of the actual retainment of the children’s bilingualism and trilingualism.

## CONCLUSION

The present analysis aimed to investigate vocabulary knowledge of trilingual migrant children in relation to the reported patterns of their language use. Crucially, we have shown that the majority language (English) of the migrant children may take care of itself, but this is not the case with the home language (Polish). We have linked these results to the relative frequency of the input in Polish and demonstrated that receiving less input in the home language may hinder vocabulary acquisition in that language.

The novelty of the paper was twofold. First, to the best of our knowledge, no previous research has investigated vocabulary acquisition in trilingual migrant children in a group study – previous papers on this topic were for the most part case studies. Second, we have contrasted trilingual language acquisition with language outcomes of bilingual and monolingual peers. We hope that this study will increase the interest in trilingual language acquisition in children and lay foundations for further investigations of the kind.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Ethics Committee at Faculty of Psychology, University of Warsaw, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee at Faculty of Psychology, University of Warsaw.

## AUTHOR CONTRIBUTIONS

Conception or design of the work: KM, MŁ, JK, AK, ZW, EH; data collection: KM, JK, AK; data analysis and interpretation: KM, MŁ, JK, AK, ZW, EH; drafting the article: KM, MŁ, JK, AK, ZW, EH; critical revision of the article: KM, MŁ, JK, AK, ZW, EH; final approval of the version to be published: KM, MŁ, JK, AK, ZW, EH.

## FUNDING

The data for this paper come from the Bi-SLI-Poland project entitled “Cognitive and language development of Polish bilingual children at the school entrance age – risks and opportunities” conducted within the European COST Action IS0804 “Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment” and carried out at the Faculty of Psychology, University of Warsaw, Poland in collaboration with Institute of Psychology, Jagiellonian University, Poland. The project was supported by the Polish Ministry of Science and Higher Education/National Science Centre (Decision 809/N-COST/2010/0). Data collection and coding were also partly supported by the Polish Ministry of Science and Higher Education grant (0094/NPRH3/H12/82/2014) carried out at the Faculty of Modern Languages, University of Warsaw, Poland, and Foundation for Polish Science subsidy to ZW.

## ACKNOWLEDGMENTS

We express our gratitude to all children and parents who participated in the study, as well as to teachers in preschools who helped in conducting the study. We thank Professor Elin Thordardottir and Professor Agnieszka Otwinowska-Kasztelanica for their helpful comments and suggestions on the previous version of the manuscript. Last but not least, we would like to thank: Professor Theo Marinis (University of Reading) and Yuly Sthefany Mora Herrera for the help in the collection of data from English monolingual children, Dr. Jakub Szweczyk for creating and sharing the script for pair-wise participant matching, and all the Bi-SLI-PL members for their unwavering commitment and considerable effort devoted to the study.

## REFERENCES

- Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism* 13, 525–531. doi: 10.1017/S1366728909990423
- Blom, E., and Bosma, E. (2016). The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian–Dutch children. *J. Child Lang.* 43, 581–607. doi: 10.1017/S0305000915000574
- Bohnacker, U., Lindgren, J., and Öztekin, B. (2016). *Turkish-and German-Speaking Bilingual 4-to-6-Year-Olds Living in Sweden: Effects of Age, SES and Home Language Input on Vocabulary Production*. Available at: <https://ir.canterbury.ac.nz/handle/10092/12907>
- Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism* 20, 3–16. doi: 10.1017/S1366728915000863
- Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., et al. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests?: language exposure and bilingual screening. *Int. J. Lang. Commun. Disord.* 49, 649–671. doi: 10.1111/1460-6984.12082
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Appl. Psycholinguist.* 28, 411–424. doi: 10.1017/S0142716407070221
- De Houwer, A. (2014). “The absolute frequency of maternal input to bilingual and monolingual children: a first comparison,” in *Input and Experience in Bilingual Development*, eds T. Grüter and J. Paradis (Amsterdam: John Benjamins). doi: 10.1075/tilar.13.03deh
- dos Santos, C., and Ferré, S. (2016). A nonword repetition task to assess bilingual children's phonology. *Lang. Acquis.* 1–14. doi: 10.1080/10489223.2016.1243692
- Dunn, L. M., Dunn, J., and Styles, B. (2009). *British Picture Vocabulary Scale*, 3rd Edn. (BPVS3). London: GL Assessment.
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., and Zebib, R. (2016). How to identify SLI in bilingual children: a study on sentence repetition in French. *Lang. Acquis.* 1–17. doi: 10.1080/10489223.2016.1192635
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language takeover. *Bilingualism* 12, 213–237. doi: 10.1017/S1366728909004015
- Gollan, T. H., Starr, J., and Ferreira, V. S. (2015). More than use it or lose it: the number of speakers effect on heritage language proficiency. *Psychon. Bull. Rev.* 22, 147–155. doi: 10.3758/s13423-014-0649-7
- Goodman, J. C., Dale, P. S., and Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *J. Child Lang.* 35, 515–531. doi: 10.1017/S0305000907008641
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459
- Grzymała-Moszczyńska, H., Grzymała-Moszczyńska, J., Durlik-Marcinowska, J., and Szydłowska, P. (2015). *(Nie)łatwe Powroty Do domu?: Funkcjonowanie Dzieci i Młodzieży Powracających z Emigracji*. Warszawa: Fundacja Centrum im. prof. Bronisława Geremka.
- Haman, E., Fronczyk, K., and Łuniewska, M. (2012). *Obrazkowy Test Słownikowy – Rozumienie (OTSR)*. Gdańsk: Pracownia Testów Psychologicznych i Pedagogicznych.
- Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*, Vol. xxiii. Baltimore, MD: Paul H Brookes Publishing.
- Hart, B., and Risley, T. R. (2003). The early catastrophe. The 30 million word gap. *Am. Educ.* 27, 4–9.
- Hoff, E. (2003). The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev.* 74, 1368–1378. doi: 10.1111/1467-8624.00612
- Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev. Psychol.* 49, 4–14. doi: 10.1037/a0027238
- Hoff, E., and Core, C. (2013). Input and language development in bilingually developing children. *Semin. Speech Lang.* 34, 215–226. doi: 10.1055/s-0033-1353448
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Hoffmann, C. (1985). Language acquisition in two trilingual children. *J. Multiling. Multicult. Dev.* 6, 479–495. doi: 10.1080/01434632.1985.9994222
- Hoffmann, C. (2001). Towards a description of trilingual competence. *Int. J. Biling.* 5, 1–17. doi: 10.1177/13670069010050010101
- Kimbrough Oller, D. (2010). All-day recordings to investigate vocabulary development: a case study of a trilingual toddler. *Commun. Disord. Q.* 31, 213–222. doi: 10.1177/1525740109358628
- Klassert, A., Gagarina, N., and Kauschke, C. (2014). Object and action naming in Russian- and German-speaking monolingual and bilingual children. *Biling. Lang. Cogn.* 17, 73–88. doi: 10.1017/S1366728913000096
- Leseman, P. P. M. (2000). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *J. Multiling. Multicult. Dev.* 21, 93–112. doi: 10.1080/01434630008666396
- Montanari, S. (2009). Pragmatic differentiation in early trilingual development. *J. Child Lang.* 36, 597. doi: 10.1017/S0305000908009112
- Montanari, S. (2010). Translation equivalents and the emergence of multiple lexicons in early trilingual development. *First Lang.* 30, 102–125. doi: 10.1177/0142723709350528
- Montrul, S. A. (2008). *Incomplete Acquisition in Bilingualism*. Amsterdam: John Benjamins. doi: 10.1075/sibil.39
- Oller, D. K., Zurer Pearson, B., and Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Appl. Psycholinguist.* 28, 191–230. doi: 10.1017/S0142716407070117
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Appl. Psycholinguist.* 18, 41–58. doi: 10.1017/S0142716400009863
- Pearson, B. Z., Fernández, S. C., and Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43, 93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x
- Rinker, T., Budde-Spengler, N., and Sachse, S. (2017). The relationship between first language (L1) and second language (L2) lexical development in young Turkish-German children. *Int. J. Biling. Educ. Biling.* 20, 218–233. doi: 10.1080/13670050.2016.1179260
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 15, 426–445. doi: 10.1177/1367006911403202
- Tuller, L. (2015). “Clinical use of parental questionnaires in multilingual contexts,” in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters).
- Unsworth, S. (2012). “Quantity-oriented and quality-oriented exposure variables in simultaneous bilingual acquisition,” in *Proceedings of the Papers of the Anêla 2012 Applied Linguistics Conference*, eds N. de Jong, K. Juffermans, M. Keijzer, and L. Rasier (Delft: Eburon), 13–22.
- Unsworth, S. (2013). Current issues in multilingual first language acquisition. *Annu. Rev. Appl. Linguist.* 33, 21–50. doi: 10.1017/S0267190513000044
- Unsworth, S. (2016). “Quantity and quality of language input in bilingual language development,” in *Lifespan Perspectives on Bilingualism*, eds E. Nicoladis and S. Montanari (Berlin: De Gruyter).
- Williams, K. (2007). *Expressive Vocabulary Test*, 2nd Edn. San Antonio, TX: Pearson Assessments.

**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 © 2017 Mieszkowska, Łuniewska, Kolak, Kacprzak, Wodniecka and Haman. 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.



# Pronoun Interpretation in the Second Language: Effects of Computational Complexity

Roumyana Slabakova<sup>1,2\*</sup>, Lydia White<sup>3</sup> and Natália Brambatti Guzzo<sup>3</sup>

<sup>1</sup> Modern Languages and Linguistics, University of Southampton, Southampton, United Kingdom, <sup>2</sup> Linguistics, University of Iowa, Iowa City, IA, United States, <sup>3</sup> Linguistics, McGill University, Montreal, QC, Canada

Children acquiring their native language (L1) have been reported to have greater difficulty in interpreting pronouns than reflexives. In addition, they are less accurate when pronouns refer to referential antecedents than to quantified antecedents, and when they hear full pronouns as opposed to reduced pronouns. We hypothesize that similar difficulties of interpretation will occur for (non-advanced) second language (L2) learners, due to an elevated computational burden, as argued for L1 acquisition by Reinhart (2006, 2011). We report on an experiment with adult learners of English (L1s French and Spanish), using a truth-value judgment task. Participants interpreted reduced and full pronouns bound by referential and quantified antecedents in aurally presented test sentences. The learners' performance is affected by type of pronoun and antecedent. When a referential antecedent is combined with a full pronoun, learners' accuracy is significantly lower. These results are in line with Reinhart's analysis of reference set computation in processing pronouns.

**Keywords:** pronoun interpretation, referential antecedents, quantified antecedents, reduced pronouns, computational complexity, Binding Principle B

## OPEN ACCESS

### Edited by:

Theo Marinis,  
University of Reading,  
United Kingdom

### Reviewed by:

Sarah Schimke,  
Universität Münster, Germany  
Wing-Yee Chow,  
University College London,  
United Kingdom

### \*Correspondence:

Roumyana Slabakova  
r.slabakova@soton.ac.uk

### Specialty section:

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

**Received:** 14 March 2017

**Accepted:** 06 July 2017

**Published:** 21 July 2017

### Citation:

Slabakova R, White L and Guzzo NB  
(2017) Pronoun Interpretation  
in the Second Language: Effects  
of Computational Complexity.  
Front. Psychol. 8:1236.  
doi: 10.3389/fpsyg.2017.01236

## INTRODUCTION

In research on second language acquisition, as in research on child language, there has been ongoing investigation of the nature of the linguistic competence achieved by learners, in the course of development as well as in the endstate. From early on, the claim has been that interlanguage grammars are systematic, conforming to the properties of natural language (e.g., Corder, 1967; Selinker, 1972; Adjémian, 1976; see also White, 2003). At the same time, it is clear that other factors may impinge, such that second language learners/speakers (henceforth L2ers) show non-native performance even when their competence can be demonstrated to be native-like. For example, there have been proposals that L2ers are not able to access full representations when parsing (the Shallow Structure Hypothesis) (Clahsen and Felser, 2006); there have been proposals that L2ers may have difficulties integrating syntactic knowledge with discourse requirements (the Interface Hypothesis) (Sorace and Filiaci, 2006; Belletti et al., 2007); there have been proposals that morphological problems exhibited by L2ers reflect difficulties in accessing forms that are in fact present in the interlanguage lexicon, possibly under production pressure when speaking (the Missing Surface Inflection Hypothesis) (Haznedar and Schwartz, 1997; Prévost and White, 2000; see also Lardiere, 2000).

In this paper, we explore another possible factor which may affect L2 performance, namely computational complexity, identified by Grodzinsky and Reinhart (1993) and Reinhart (2006, 2011) as accounting for L1 acquirers' relatively poor performance in interpreting referents for

pronouns in certain contexts, compared to their performance on reflexives. We will show that L2ers have a problem with pronoun reference which is similar to (though not as severe as) child L1 acquirers; we suggest that the reason is the same, namely the computational complexity of the structure in question. This complexity may translate into an elevated processing load, though this is not directly tested in our study.

In order to explore this issue, we investigate the so-called Delay of Principle B Effect (DPBE) in adult learners of English. Principle B of the Binding Theory (Chomsky, 1981) constrains the distribution of pronouns (see below). Research on the DPBE in L1 acquisition has shown that children do not suffer from a representational deficit: Principle B is in fact present in child grammar but other factors sometimes cause children to fail to observe this principle. We suggest that the same holds true in L2 acquisition, at least in the case of learners who are not of advanced proficiency.

## PRINCIPLE B AND THE DPBE IN CHILD LANGUAGE

Pronouns (*him*, *her*, etc.) behave differently from anaphors like reflexives (*himself*, *herself*, etc.). In the typical case, the antecedent of an anaphor cannot occur in the same position as the antecedent of a pronoun.<sup>1</sup> In particular, anaphors require their antecedents to be close (or *local*) whereas pronouns disallow this. Consider the English examples in (1) and (2):

- (1) Mary<sub>i</sub> thought [that Susan<sub>j</sub> liked herself<sub>j/\*i</sub>]
- (2) Mary<sub>i</sub> says [that Susan<sub>j</sub> likes her<sub>i/k/\*j</sub>]

In (1), the reflexive *herself* can only refer to the local antecedent, *Susan*, and not to the non-local antecedent, *Mary*. In (2), on the other hand, *Susan* is impossible as an antecedent for the pronoun, whereas *Mary* (or anyone else of female gender mentioned in the previous discourse) is a possible antecedent.

To express these relationships, Chomsky (1981) formulated Principles A and B of the Binding Theory, presented, in simplified form, below, where *local* means roughly “in the same clause”:

- (3) Principle A: a reflexive must take a local antecedent.
- (4) Principle B: a pronoun may not take a local antecedent.

In other words, Principle B renders local antecedents ‘inaccessible’ to pronouns.

It turns out that acquisition of pronouns, particularly with respect to choice of antecedents, presents rather distinctive challenges for children acquiring their first language (L1). In the acquisition of English and many other languages, a well-known and robust phenomenon known as the DPBE has been reported (Jakubowicz, 1984; Crain and McKee, 1985; Chien and Wexler, 1990; Koster, 1993; Avrutin and Thornton, 1994; Thornton and Wexler, 1999; among many others). In a nutshell, children are often at chance when interpreting sentences with pronouns, at stages when they have no problem in interpreting reflexives. In

particular, they sometimes mistakenly assume that pronouns, like reflexives, can take local antecedents.

Delays in acquiring accuracy on pronouns have been observed cross-linguistically, for Dutch (Philip and Coopmans, 1996), Hebrew (Friedmann et al., 2010), Icelandic (Sigurjónsdóttir, 1992) and Russian (Avrutin and Wexler, 1992) but not for languages which have clitic pronouns (Spanish: Baauw et al., 1997; Baauw, 2002; Baauw and Cueto, 2003; French: Zesiger et al., 2010; Italian: McKee, 1992; Greek: Varlokosta, 2002).

There is a further relevant finding in the literature, relating to whether the antecedent is referential (referring to a particular individual, e.g., *Mama Bear*) or quantificational (referring to some quantified group, e.g., *every bear*). Chien and Wexler (1990) found that 6-year-old children were much more accurate with quantified antecedents than with referential, mostly rejecting local antecedents for pronouns in the former case (84% rejection) while rejecting them in the latter only around 50% of the time. This finding has come to be known as “the quantificational asymmetry” in the interpretation of pronouns.

More recently, an additional asymmetry has been reported. Hartman et al. (2012) compared performance on fully pronounced versus phonologically reduced pronouns with referential antecedents, such as (5).

- (5) I think... Cow washed 'm.

Hartman et al. (2012) used a truth-value judgment task (TVJT), in which participants saw stories acted out with toys, each story being paired once with a full pronoun test sentence and once with a reduced pronoun. In their experiment, children's correct rejections of local antecedents for full pronouns were around 53% (similar to findings by Chien and Wexler, amongst others); on the other hand, rejection of local antecedents for reduced pronouns was significantly higher, at 80.6%.

To summarize so far, child language research has established that there is a DPBE in children's comprehension. However, children are less accurate with pronouns referring to referential antecedents than with pronouns where the antecedent is quantified. Furthermore, a full pronoun versus reduced/clitic pronoun asymmetry is attested. Accuracy with quantified antecedents and with reduced pronouns suggests that Principle B is indeed operative and that some other explanation is required to account for the problematic cases.

## Toward an Explanation: Accidental Coreference

We turn now to an explanation of why pronoun reference should be particularly difficult to acquire, proposed by Grodzinsky and Reinhart (1993) and Reinhart (2006, 2011). There are two ways in which a pronoun and its antecedent can be associated. In addition to variable binding of pronouns (as regulated by Principle B), accidental coreference is also possible (Sag, 1976; Evans, 1980; Grodzinsky and Reinhart, 1993; Heim, 1993; Williams, 1977). In very specific contexts, a pronoun can in fact take a local

<sup>1</sup>We exclude from consideration here anaphors which allow long-distance antecedents, as found in languages like Chinese, Japanese, or Korean, for example.

antecedent. Such cases are heavily dependent on repetition and special intonation.

- (6) A: Is that speaker Zelda?  
B: She must be. She praises her to the sky.
- (7) From the movie “Side effects” (2013), directed by Steven Soderbergh, spoken by a psychiatrist to explain an accident with a patient:

“The patient blamed me. The patient’s wife blamed me. The patient’s children blamed me. Even I blamed me.”

These examples ostensibly violate Principle B, since the pronoun and its antecedent are in the same clause. Linguists have dealt with this problem by assuming that different indices have in fact been assigned to the pronoun and the antecedent; they just happen to refer to the same person, as shown in (6B’):

- (6B’) She<sub>i</sub> must be. She<sub>i</sub> praises her<sub>j</sub> to the sky. where  $i = j$  accidentally.

The assumption, then, is that, in interpreting pronouns, two derivations have to be constructed and compared. Reinhart (2006, 2011) calls this phenomenon “reference set computation” and invokes it as an explanation of other linguistic phenomena, such as Focus and scalar implicatures. Pronoun interpretation is computationally more complex than anaphor computation, for which only one interpretive mechanism exists, namely variable binding. As far as child language is concerned, Grodzinsky and Reinhart (1993) and Reinhart (2006, 2011) argue that the necessity for reference set computation with pronouns taxes children’s working memory resources; more specifically, reference set computation “relies heavily on the ability to store and perform further computation on temporary outcomes” (Reinhart, 2011: 168). On this account, when trying to interpret pronouns, children sometimes give up and pick an interpretation at random. This difference in computational complexity accounts for children’s roughly 50% accuracy on pronouns and their superior accuracy on reflexives.

This account also explains children’s accuracy with quantified antecedents, since these are subject only to variable binding, no accidental coreference being possible in such cases. The account has also been used to explain children’s accurate performance on pronouns in languages with clitics. According to Avrutin and Wexler (1992), accidental coreference is unavailable with clitic pronouns, because clitics are referentially deficient, in the sense that they are always bound variables (see also Baauw and Cuetos, 2003). This is suggested by the fact that clitics cannot be used in isolation, cannot receive focal stress, and cannot be used deictically with a pointing gesture. Children learning languages with clitics do not consider an accidental coreference derivation because of the requirement that the clitic is always coindexed with its antecedent, and so they are more accurate than children learning languages with strong pronouns, which are free to take on accidental coreference. English phonologically reduced pronouns, such as *’m* for *him*, are similar to clitics in this respect.

To summarize so far, children’s greater success with quantified antecedents and reduced pronouns can be explained if children

engage in reference set computation (deciding between binding and accidental coreference) only with full pronouns and with referential antecedents. In other words, computational complexity rather than lack of linguistic knowledge is the source of their difficulties. The account is potentially extendable to adult L2 learners.

## PRONOUN INTERPRETATION IN L2 ACQUISITION

The issue of potential computational complexity, as defined by Reinhart and colleagues with respect to pronouns, has not been addressed in L2 acquisition. While there are a number of studies on the L2 acquisition of reflexives and their antecedents, less is known about pronouns. If the difference in accuracy in determining antecedents for pronouns and reflexives is computationally based, it is logical to assume that the same dissociation between pronouns and anaphors may arise in L2 acquisition as well. However, additional factors come into play in adult L2 acquisition. First, adult learners, having a fully developed computational system for their L1, may not display a big contrast between pronoun and anaphor interpretation accuracy in the L2, because they have learned to compute these meanings as children in their L1. Second, all languages have personal pronouns, in some cases taking the form of clitics, so L1 transfer into the L2 is possible, including transfer of requirements on possible antecedents. These two factors could aid learners in acquiring pronoun reference, and may obscure any computational effects that arise in the course of acquisition. Indeed, in the past, the understanding was that there are no significant problems with pronoun interpretation in L2 acquisition as far as Principle B is concerned (White, 1998).

Nevertheless, a new look at this phenomenon is warranted. First, the predictions made by the computational complexity account extend naturally to lower proficiency L2 learners, who may exhibit greater signs of struggling with pronoun interpretation than more experienced learners. Furthermore, new research using psycholinguistic techniques such as eye tracking (Kim et al., 2015) has already suggested that the processing of pronouns differs from the processing of anaphors, at least for Korean speakers of L2 English.

We turn now to a summary of previous research on pronoun interpretation in L2 as it relates to Binding Principles A and B. There has been extensive research on Principle A, looking at properties of reflexive pronouns, and focusing in particular on cross-linguistic differences that might come into play when the L1 and L2 differ with respect to whether long-distance antecedents are permitted (e.g., Finer and Broselow, 1986; Hirakawa, 1990; Thomas, 1995). There has been less work on Principle B. A few studies are relevant, either implicitly or explicitly, to the question of whether or not there is a DPBE in L2 acquisition; in particular, there are studies that compare performance on Principles A and B, looking only at cases involving referential antecedents.

Finer and Broselow (1986) were among the first to look at acquisition of an L2 (English) which permits only local antecedents for reflexives by speakers whose L1 (Korean) permits

long-distance antecedents. Results from their pilot study of Korean learners of English ( $n = 6$ ) on reflexives are well known: in tensed clauses, only local antecedents for reflexives were accepted, whereas in non-finite clauses non-local antecedents were accepted 40% of the time. What is less well known is that this study also included an examination of pronouns with referential antecedents. Results show that this small group of L2 learners accepted local antecedents for pronouns 46% of the time in tensed clauses and 21% of the time in non-finite clauses.<sup>2</sup> In other words, if we consider only tensed clauses, they were much more accurate on interpretation of reflexives than pronouns, suggesting (indirectly) a possible DPBE.

Lee and Schachter (1997) argue for windows of opportunity in L2 acquisition, proposing that there are sensitive periods for L1 and L2 acquisition, such that certain properties cannot be successfully acquired before the onset of the sensitive period or after the end of it. Lee and Schachter tested this claim by looking at the L2 acquisition of Binding Principles A and B by Korean-speaking learners of English, with different ages of onset for the acquisition of English. Participants were tested on properties of reflexives and pronouns by means of a TVJT. Learners fell into various age categories at time of testing. The youngest groups (6–7 and 8–10 year olds) performed better on reflexives than on pronouns, consistent with the idea that the windows of opportunity open at different times for these two principles, and also consistent with a DPBE.

White (1998) investigated pronoun interpretation by Japanese-speaking and French-speaking learners of English, of high intermediate proficiency, hypothesizing that adult learners would not show problems with pronouns, on the assumption that difficulties with pragmatics, processing or computation, argued to account for the difficulties of children, would not arise for adults. Results from a TVJT show that the L2 groups appropriately rejected local antecedents for pronouns. In other words, there was no evidence of a DPBE in the groups as a whole. However, there were three participants (out of 28), one francophone and two Japanese speakers, who consistently accepted local antecedents for pronouns.

Two recent studies investigated anaphor and pronoun interpretation in L2 acquisition using eye tracking. Patterson et al. (2014) tested advanced German-speaking learners of English, to determine whether they know that a local antecedent for a pronoun is ‘inaccessible’ according to Principle B. In their experiment 2, participants read sentences which manipulated the gender of the potential antecedents. Native speakers and L2ers behaved alike: the non-local mismatch condition (sentences like *Jane remembered that John had taught him a new song*) resulted in longer reading times than the other conditions (*John remembered that Jane had taught him a new song*; *John remembered that Mark had taught him a new song*). While such results are consistent with the claim that L2ers are observing Principle B, the researchers question this interpretation. They added another experiment, involving clauses containing prepositional phrases (e.g., *Barry saw Gavin place a gun near him*). In such cases, the pronoun

exceptionally allows a local antecedent (here *Gavin*), in violation of Principle B. Native speakers showed longer reading times when the object mismatched the pronoun in gender (e.g., *Barry saw Megan place a gun near him*), suggesting they were expecting a local antecedent for the pronoun. The L2ers, in contrast, showed longer reading times when the pronoun and the subject mismatched (e.g., *Megan saw Barry place a gun near him*). The researchers attribute the L2ers’ results not to Principle B but to “a general preference to link the pronoun to the matrix subject” (p. 15), and suggest that this also explains their success in experiment 2. We return to this issue in the discussion.

The second study to use eye-tracking, Kim et al. (2015), compared performance on Principles A and B. Assuming the Reflexivity Theory approach to binding (Reinhart and Reuland, 1993; Reuland, 2001, 2011), Kim et al. (2015) predicted that reflexives, being licensed syntactically, would be easier to interpret than pronouns, which in this framework require access to a pragmatic module in addition to syntax. The study used the visual world paradigm. Participants were adult native speakers of English as well as Korean-speaking learners of English, of intermediate to advanced proficiency.

Participants had to manipulate various cartoon characters displayed on the screen, in accordance with auditory instructions. With a mouse click, a character could be picked up and moved along a trajectory to a goal. Results were calculated in terms of the correct movement of the characters toward a potential antecedent as well as by the speed of eye fixation onto the place where the character had to be moved. Results indicate that when they heard a sentence with a pronoun such as *Look at Goofy. Have Mickey touch him*, the native speakers overwhelmingly chose the antecedent to be Goofy. The learners also predominantly chose Goofy as the antecedent; however, they also incorrectly chose Mickey as a possible antecedent 24% of the time, suggesting a DPBE effect, since they were totally accurate in the case of reflexives.

Furthermore, comparing the time it took the participants to start looking at the subject of the test sentence when they heard the lead-in sentences, the native speakers looked at the subject character (Mickey) no more in the pronoun condition than in the name condition (*Have Mickey touch Donald*). The L2 learners, however, looked at Mickey significantly more in the pronoun condition, suggesting that they were considering Mickey as a potential antecedent. There was also a proficiency effect, in the sense that the lower proficiency learners took much more time to resolve the antecedent issue. The researchers concluded that the learners interpreted reflexives in a nativelike way, but demonstrated much more inaccuracy, hesitation and time delays when processing pronouns.

Few L2 studies have compared performance on referential and quantificational antecedents. One exception is Marinis and Chondrogianni (2011) who investigated the comprehension of reflexives and pronouns by children who are sequential bilinguals (L1 Turkish, L2 English). These children (mean age 7.8, ranging from 6.2 to 9.9) were compared to L1 acquirers of English (mean age 7.5, ranging from 6.0 to 9.0). The task, once again, was a TVJT. Test items included reflexives

<sup>2</sup>Finer and Broselow do not, in fact, discuss their results relating to pronouns, but they are available in an appendix.

and pronouns; antecedents were referential or quantificational. While Marinis and Chondrogianni do not directly compare performance on reflexives with performance on pronouns, they do show that the bilingual children performed like the monolinguals on reflexives and were less accurate than monolinguals on pronouns, which suggests that Principle B was more problematic for them than Principle A. Both groups showed a quantificational asymmetry in the case of pronouns.

Before turning to our own study, we briefly mention a different kind of approach, namely the Interface Hypothesis (Sorace and Filiaci, 2006), which also predicts problems with pronoun interpretation in L2. Sorace and Filiaci (2006) and Belletti et al. (2007) report that advanced and near-native speakers of L2 Italian occasionally overuse overt subject pronouns in contexts where null pronouns would be preferred by native speakers. They attribute this overuse to problems at the syntax-discourse interface, namely a failure to fully appreciate the discourse requirements on overt pronouns, which imply a change in topic, unlike null pronouns which indicate topic continuity. The work of these researchers has focused on interpretation of subject pronouns, where Principle B is not at issue. Nevertheless, there are some commonalities in that processing problems have been suggested as an explanation (Sorace, 2011, 2016), a point we return to in the discussion.

The research described above suggests that all might not be well when it comes to pronoun interpretation in the second language. In the following section, we report on an experiment to investigate whether or not there is a DPBE effect in L2 and, if so, whether it is attributable to computational complexity. Our experiment does not focus on the comparison between anaphors and pronouns but instead on the interpretation of reduced versus full pronouns, and on the quantificational asymmetry with full pronoun antecedents. To anticipate the findings, we will show that learners of L2 English experience difficulties with pronoun interpretation. However, this only happens when full pronouns are combined with referential antecedents. In addition, learners' interpretations are constrained by their level of proficiency in English. These findings are consistent with the assumption that computational complexity of the kind envisaged by Reinhart and colleagues is implicated.

## THE PRESENT STUDY

### Predictions

In section "Principle B and the DPBE in Child Language," we presented the well-known delay in the correct interpretation of pronouns by children. As already discussed, we follow Grodzinsky and Reinhart (1993) and Reinhart (2006, 2011) in assuming that the DPBE reflects difficulties due to computational complexity caused by having to determine whether or not accidental coreference comes into play. We expect a similar difficulty of interpretation for L2ers, at least at lower levels of proficiency, attributable to the need to compute accidental coreference in the L2. Since accidental coreference is not

possible with reduced pronouns or with quantified antecedents, we predict that learners will have difficulties only in cases where a full pronoun takes a referential antecedent. To investigate this prediction, we set out to establish whether learners of English with French or Spanish as their native languages correctly interpret sentences with reduced and full pronouns bound by referential and quantificational antecedents.

As discussed above, English is a language which has both strong and weak (phonologically reduced) forms of object pronouns (such as *him* versus *'m*). In contrast, French and Spanish are languages with object clitic pronouns, which differ in a number of respects from strong pronouns (see Kayne, 1975, for French). For example, as mentioned above, clitics cannot occur in isolation and are unstressed. They also differ from strong pronouns in their syntactic positions: object clitics are preverbal when the verb is finite. Spanish and French differ somewhat with respect to placement of clitics with non-finite verbs. We put these differences to one side as our test items only include finite verbs and the position of object pronouns is not under investigation. Given the similarities between French and Spanish with respect to object clitics, we do not expect differences in response patterns based on L1.

### Participants

A hundred and twenty-five individuals participated in two experiments: 65 in the Full Pronoun experiment and 60 in the Reduced Pronoun experiment. They comprised two groups of English native speakers, mostly recruited in Montreal, QC, Canada, and Southampton, United Kingdom, and four groups of learners of English with French or Spanish as their native languages, recruited and tested in Montreal. See **Table 1** for details.

The learners in both experiments had similar profiles. Most of them reported that they started learning English in a school setting (82.6%). The average age at which learners started to acquire English was 11.2, most of them between the ages 10 and 18 (60.5%). The majority of the learners were living in Montreal, QC, Canada, for work or study purposes. Some indicated having some knowledge of other languages (including French in the case of the native speakers of Spanish). Seven learners reported that they were taking English classes at the time of their participation in the experiment.

Testing took place individually (or in small groups in the case of native speakers) in a quiet lab. Participants took about half an hour to do the test (plus about 10 min for the proficiency test, in the case of the learners) and were remunerated for their participation.<sup>3</sup>

### Proficiency Test

Learners' proficiency in English was assessed through an adapted version of the Oxford Test of Proficiency. The test included 40

<sup>3</sup>Informed consent was obtained from all participants. The research program under which this project was conducted was reviewed by Research Ethics Board II of McGill University and is deemed to be in compliance with the ethical standards expected for research with human subjects (approvals: REB #451-0511 and #60-0715).

**TABLE 1 |** Participants in the two experiments.

|                        | Full pronoun experiment |        |          | Reduced pronoun experiment |        |          |
|------------------------|-------------------------|--------|----------|----------------------------|--------|----------|
|                        | <i>n</i>                | Female | Mean age | <i>n</i>                   | Female | Mean age |
| Native speakers        | 20                      | 11     | 26.7     | 19                         | 9      | 28.9     |
| French-speaking L2ers  | 28                      | 18     | 27.7     | 22                         | 15     | 34.5     |
| Spanish-speaking L2ers | 17                      | 10     | 28.7     | 19                         | 10     | 32.5     |

grammar-based multiple-choice items, with a maximum score of 40. Learners' mean proficiency scores for the two experiments are similar: 29.1 for the reduced pronoun experiment (range: 17–39), and 29.2 for the full pronoun experiment (range: 13–39). As will be discussed in the next section, we treat proficiency as a continuous variable.

### Truth Value Judgment Task (TVJT)

The TVJT (Crain and McKee, 1985; Gordon, 1996; Crain and Thornton, 1998) tests a speaker's ability to evaluate interpretations of test sentences in controlled contexts/scenarios. The participant must decide whether a test statement is True or False as a description of a particular situation. A fundamental requirement of such tasks (Crain and Thornton, 1998) is that the story renders a grammatical reading false; consequently, only responses to stimuli expecting the answer False are considered to be truly informative of participants' underlying grammatical competence. Furthermore, there is a Condition of Plausible Dissent (Crain and Thornton, 1998) or a Disputability Requirement (Conroy et al., 2009). The Condition of Plausible Dissent is satisfied if the grammatically inaccessible antecedent has been under consideration and is a genuine potential outcome of the story that almost comes to pass but in the end does not. This requirement ensures that the decision in the TVJT is taken on the basis of grammar, rather than the pragmatics of the story.

There is a further requirement, specific to TVJTs probing pronoun interpretation (Elbourne, 2005; Conroy et al., 2009): the Availability Requirement. Elbourne (2005) critiqued previous experiments for not making the antecedent sufficiently prominent in the story's discourse. Only if children reject an available and prominent antecedent can we be certain that it is the child's grammar, and not the discourse context, that is responsible for the attested interpretation. Following Conroy et al. (2009), we make sure this requirement is obeyed by including stories which mention groups of characters that are performing both reflexive and transitive actions. Our TVJT conforms to Conroy et al.'s (2009) recommendation that all characters mentioned in the story are sufficiently individuated to be considered as possible referents. In addition, all stories mentioned multiple characters so that the stories in the quantified antecedent condition did not involve more characters than stories in the referential antecedent condition. In the test conditions, each story is compatible with a reflexive as well as a pronominal interpretation.

In addition, we introduced another variable in our design. Within each condition (Referential antecedent, Quantified antecedent, filler), 4 sentences expected a True answer and 4 a False answer. Only the False-answer test sentences obey the

above-mentioned TVJT design requirements; those expecting the True-answer serve as additional fillers.

We did not vary the factor quantified versus referential antecedent within items, because it was difficult to construct plausible stories that would fit both types of antecedents. We also did not vary the factor reduced versus full pronouns within participants, because we were concerned that a response bias or confusion might have been introduced if learners were exposed to both types of pronouns.

In what follows, we examine some representative context stories and explain how they satisfy or fail to satisfy the Requirements of Disputability and Availability. It is important to keep in mind that the contexts were presented visually in writing (on a computer screen) and aurally; test sentences were presented only aurally, since it was crucial that participants heard the form of the pronoun (full or reduced), rather than reading it.<sup>4</sup> Each story was followed by a test sentence with either a reduced pronoun or a full pronoun, depending on the experiment.

A referential condition story with an expected False answer is exemplified in (8).

- (8) Example from the referential condition with the expected answer 'False'.

Tom, Helen, and Harry were going to a soccer party. Prizes were going to be given out for the best spray-painted logo. They all sprayed the logo of their favorite soccer teams on their arms. Tom badly wanted to win the competition, so he asked his friends to help him make his logo even better. Helen refused to help because she wanted to win as well. Harry wanted to help Tom, but he had no spray-paint left.

*Harry sprayed 'm.* (Reduced pronoun experiment) T F

*Harry sprayed him.* (Full pronoun experiment) T F

The anaphoric (local, co-referential) reading (Harry sprayed himself) is available in this story, because all the three characters sprayed the logo of their favorite teams on themselves. The non-coreferential (non-anaphoric) interpretation (Harry sprayed Tom) is potentially available and under consideration, but in the end does not come to pass because there is no paint left. Thus the requirement of Disputability is satisfied.

In order to consider the requirement of Availability further, we compare this referential condition story with a quantificational condition story such as the one in (9), in which the expected answer is also False.

- (9) Example from the quantificational condition with the expected answer 'False'.

<sup>4</sup>There were no pictures accompanying the text.

Jim, Jack, and Bert always drive to college, each of them using his own car. Their friend John doesn't own a car so Jim, Jack, and Bert all agreed to drive him to school. But this week, on Monday Jim forgot to pick John up. On Tuesday, Jack overslept and drove to class alone. Only Bert was true to his word and drove John to school on Wednesday.

*This week, every student drove 'm to school* (Reduced pronoun experiment) T F

*This week, every student drove him to school.* (Full pronoun experiment) T F

In parallel with the test item in (8), the anaphoric interpretation in (9) is available and prominent, because the three characters, Jim, Jack and Bert, always drive to school, each one using their own car, hence they drive themselves. The non-anaphoric interpretation (*Every student drove John to school*) is potentially under consideration and actually promised, but it never comes to pass due to highly individuated circumstances. Finally, the available propositions evaluated by the participants are closely matched in the stories in (8) and (9).

Let us now consider a True-answer story from the referential condition as in (10).

- (10) Example from the referential condition with the expected answer 'True'.

Christopher, Mary, and Ben work in a bakery. Christopher and Mary bake bread and pastries and Ben sells them. Mary always wears an apron but Christopher does not. At the end of each day, Christopher is very dusty from all the flour. Ben dusts his friend's clothes and hair off until Christopher is completely clean.

*Ben dusts 'm off.* (Reduced pronoun experiment) T F

*Ben dusts him off.* (Full pronoun experiment) T F

In this story, the anaphoric interpretation is missing: Ben never dusts himself off. The requirement of Disputability is also not obeyed: there is nothing to dispute since the action is actually confirmed. In addition to violating the TVJT requirements, these stories are easier to interpret, since the correct pronominal interpretation (the non-anaphoric one) is rather prominent. In addition to stories with referential or quantificational antecedents, the experiment included stories followed by test sentences containing full NPs in object position. These items were also treated as fillers; see (11).

- (11) Example of filler story with a full NP in object position in the test sentence.

Anne, Margo, Celia and Rita find an old empty house and spend all day playing inside. They get covered in dust. They try to clean the dust off themselves but Anne is no good at it. Anne asks Rita to help her, but Rita is too tired. Celia has already gone home. In the end, Margo agrees to help and does a great job.

*Margo cleans Anne.* T F

To summarize, we have 8 test items in each experiment (responses where the expected answer is False), and 16 filler.<sup>5</sup> In other words, each experiment (reduced or full pronouns) comprised 24 story-test sentence combinations: 8 test items expecting False answers, 8 fillers expecting True answers and 8 fillers with full NPs in object position, with answers that were true or false. Within the items involving pronouns, 4 had referential and 4 had quantificational antecedents. The context stories were identical in the two experiments. Test items differed, involving the full pronoun *him* in one experiment and the reduced pronoun *'m* in the other.<sup>6</sup> Each participant was tested on all 24 story-sentence combinations within one experiment; no participant undertook both experiments. The presentation software (SurveyGizmo) randomized the order of item presentation for each participant.

## Statistical Analysis

We modeled learners' responses for the target test items using a multilevel logistic regression with random effects (`glmer()` in R; R Development Core Team, 2017). The maximal converging model included the following predictors: *native language* (French or Spanish), *proficiency score* (continuous variable), *antecedent* (referential or quantified), *pronoun* (full or reduced), and the interaction between *antecedent* and *pronoun*.<sup>7</sup> We included this interaction given the hypothesis that any inaccuracy will be the result of computational complexity and is, therefore, dependent on both type of antecedent and type of pronoun. In addition, the model included a by-item random intercept and a by-speaker random slope for antecedent, to account for the variation among test items and the variation among speakers with regard to antecedent, respectively.

A separate logistic regression with the same predictors (both main effects and random effects) was run to verify whether learners' responses to the True answer fillers were affected by any of the predictors included in the analysis. In order to compare the accuracy of learners and native controls, we performed two chi-square tests, one comparing the groups with respect to their accuracy on the fillers, the other comparing their accuracy on the target items.

## Results

Participants either took part in the experiment that included reduced pronouns or the experiment that included full pronouns. As described above, the target items were those for which False answers were expected, with two types of fillers: items for which True answers were expected and items containing full NPs instead of pronouns in object position. Participants' accuracy on both types of fillers was high (Table 2); controls were more accurate than learners on fillers ( $\chi^2 = 23.1, p < 0.0001$ ).

<sup>5</sup>Hartman et al. (2012) have 4 test items expecting the answer False with full pronouns and 4 with reduced pronouns; they only tested referential antecedents.

<sup>6</sup>Female characters were introduced in the stories but did not occur in the test items.

<sup>7</sup>The underlined levels in parentheses are the reference levels for our predictors. Except for *native language*, where we do not expect to find any main effects, the reference levels for *antecedent* and *pronoun* were decided upon based on our predictions (see Predictions).

**TABLE 2 |** Mean accuracy (%) on fillers by group and experiment (reduced or full pronouns).

|              | Controls         |               | Learners         |               |
|--------------|------------------|---------------|------------------|---------------|
|              | Reduced pronouns | Full pronouns | Reduced pronouns | Full pronouns |
| Full NPs     | 98.6%            | 98.6%         | 93.2%            | 94.6%         |
| True answers | 92.05%           | 95.75%        | 82.85%           | 87.2%         |

**TABLE 3 |** Mean accuracy scores (in %) by group, pronoun, and antecedent.

|                         | Controls         |               | Learners         |               |
|-------------------------|------------------|---------------|------------------|---------------|
|                         | Reduced pronouns | Full pronouns | Reduced pronouns | Full pronouns |
| Quantified antecedents  | 100%             | 95.8%         | 87.1%            | 90.5%         |
| Referential antecedents | 98.6%            | 97.2%         | 91.4%            | 82.7%         |

The logistic regression for the True answer fillers indicates that learners' accuracy is not conditioned by antecedent, pronoun, or the interaction between antecedent and pronoun ( $p > 0.05$ ).

We now turn to participants' accuracy on the target items. **Table 3** shows mean accuracy scores on items expecting False as the answer, by pronoun and antecedent.

**Table 3** shows (a) that the controls perform at ceiling while the L2ers are in general less accurate than controls ( $\chi^2 = 24.2$ ,  $p < 0.001$ ); (b) in the case of full pronouns, the L2ers are more accurate with quantified antecedents than with referential antecedents; and (c) the L2ers are the least accurate with full pronouns taking referential antecedents. Thus, while the controls' performance is not affected by type of pronoun or antecedent, the L2ers' performance is.

**Table 4** shows the estimates of our statistical model for L2ers' performance on the target items. A positive estimate ( $\hat{\beta}$ ) indicates that the predictor in question is associated with an increase in accuracy.

The results for *native language* indicate that, as expected, there is no significant difference between French-speaking and Spanish-speaking L2ers' responses. On the other hand, learners' performance improves significantly as their scores in the proficiency test increase. Each unit<sup>8</sup> increase in proficiency test scores raises the odds of getting a right answer by a factor of 1.39 [ $\exp(\hat{\beta})$ ].

<sup>8</sup>Given that the predictor *proficiency result* was scaled and centered, each unit here is equivalent to one standard deviation in the proficiency score ( $SD = 6.58$ ).

**TABLE 4 |** Coefficient values, standard error (SE), z-value (Wald test), and p-value for predictors in the statistical model.

| Predictors                            | Estimate ( $\hat{\beta}$ ) | SE   | z-value | p-value |
|---------------------------------------|----------------------------|------|---------|---------|
| Intercept                             | 3.36                       | 0.65 | 5.1     | < 0.001 |
| Native language (French)              | -0.49                      | 0.51 | -0.96   | 0.33    |
| Proficiency score                     | 1.18                       | 0.27 | 4.3     | < 0.001 |
| Full pronoun                          | 0.3                        | 0.53 | 0.57    | 0.56    |
| Referential antecedent                | 1.04                       | 0.78 | 1.33    | 0.18    |
| Referential antecedent * Full pronoun | -1.76                      | 0.64 | -2.74   | 0.006   |

**Figures 1, 2** show the L2ers' mean accuracy on each of the four possible combinations of antecedent and pronoun. In each figure, the x-axis shows scores on the proficiency test while the y-axis shows learners' mean accuracy in the task. The darker circles indicate a higher concentration of L2ers with a given mean accuracy and proficiency score. There are two patterns of note in these figures: (a) L2ers with a lower score on the proficiency test overall perform worse than learners with a higher score on the proficiency test, and (b) the combination of a referential antecedent and a full pronoun (left panel in **Figure 2**) yields a higher concentration of lower scores than the other possible combinations between antecedent and pronoun in the data, as indicated by the steeper slope of the trend line. In particular, problems do not arise with quantified antecedents or with reduced pronouns.

The model indicates that the interaction between *antecedent* and *pronoun* is significant: when a referential antecedent is combined with a full pronoun, learners' accuracy goes down, as suggested by the trend lines in **Figure 2** and the values in **Table 3**. This interaction has a negative effect on L2ers' accuracy, which is consistent with our hypothesis. *Pronoun* and *antecedent*, however, are not significant as main effects.

In summary, L2ers' accuracy on test items is affected by *proficiency score* and the interaction between *antecedent* and *pronoun*: learners who are more proficient are overall more accurate, and learners' performance is worse on the combination between referential antecedents and full pronouns. The next section discusses these results in light of our predictions.

## DISCUSSION

Let us recap the predictions and significant findings of this study. We set out to evaluate pronoun reference by native speakers and L2ers, in the light of difficulties exhibited by L1 acquirers, the so-called DPBE. We evaluated L2ers' interpretations in two experiments, identical except for the form of the pronoun: in one, participants heard full pronouns in the test sentences; in the other, they heard phonologically reduced pronouns. In both cases, test items involved quantificational and referential antecedents. Participants had to evaluate the

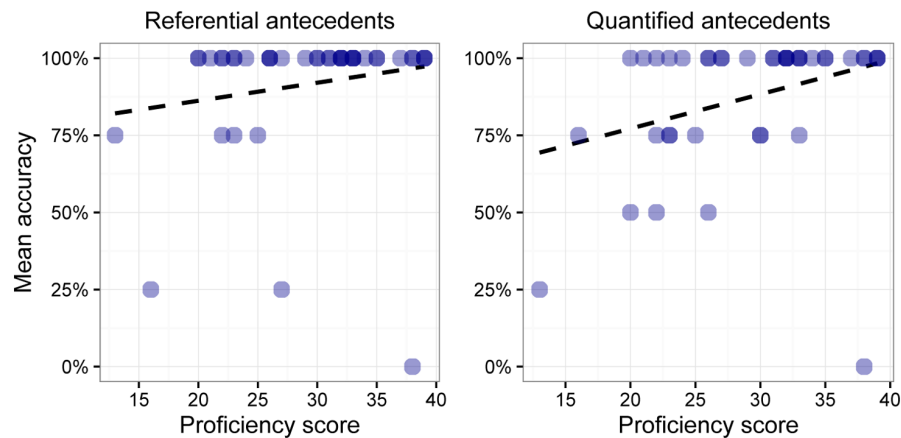


**FIGURE 1** | Individual accuracy with reduced pronouns.

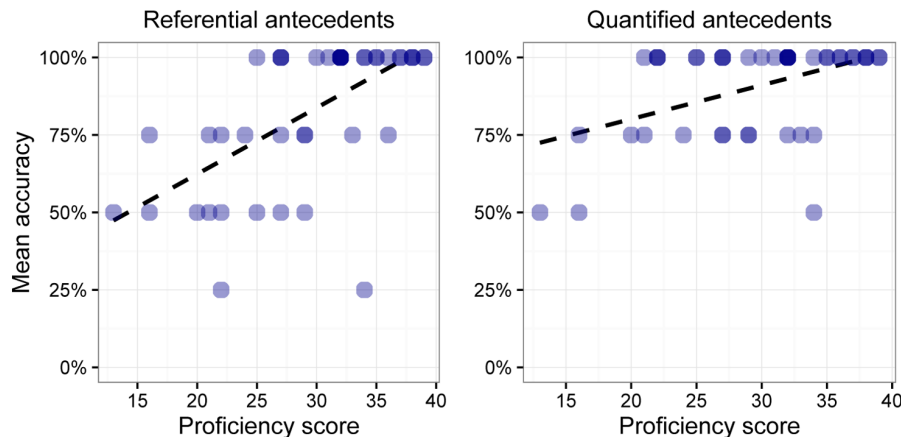


**FIGURE 2** | Individual accuracy with full pronouns.

truth of the test sentences in a TVJT with contexts presented in written and spoken form, and test sentences presented only aurally.

As far as we are aware, no study of L1 or L2 acquisition has looked at the combination that we investigated, namely a comparison of referential and quantified antecedents for full and reduced pronouns. This combination is essential to fully assess the potential role of reference set computation in learners' determination of antecedents for pronouns. Given findings in the L1 literature that children have greater difficulties with pronouns with referential antecedents than with quantified antecedents, the so-called quantificational asymmetry, and greater difficulties with full pronouns than with reduced pronouns, we expected to find lower accuracy on referential antecedents but only in the case of full pronouns. This prediction was supported by the multilevel logistic regression results reported in the previous section.

As far as the lower proficiency L2ers are concerned, we observed greater accuracy on quantified antecedents than on referential antecedents with full pronouns, as can be seen in **Figure 2**. We also established greater accuracy with reduced

pronouns versus full pronouns, in the case of referential antecedents; see **Table 3** and the left-hand panels of **Figures 1, 2**. Lower proficiency L2ers achieved relatively high accuracy in the reduced pronoun version of the experiment (**Figure 1**). The steeper slope on the left panel of **Figure 2** indicates that the learners are less accurate in the full pronoun-referential antecedent combination.

More advanced learners did not exhibit a quantificational asymmetry, nor did they manifest reduced accuracy with full pronouns, as can be verified by looking at the higher proficiency individuals in **Figures 1, 2**. They were able to identify the correct antecedents for all pronouns in each experiment. The same pattern was observed in the native speakers; see **Table 2**. These findings suggest that advanced and native speakers were essentially performing at ceiling.

Our findings are easily accounted for in terms of the computational complexity proposal of Grodzinsky and Reinhart (1993) and Reinhart (2006, 2011). These researchers argue that when the antecedent is a referential NP, children have to consider both variable binding as well as accidental coreference as

possible routes to finding an appropriate referent. Constructing the reference set, keeping it in short-term memory and comparing the two derivations proves costly, and in the end children give up and choose an available antecedent at random. Quantificational antecedents, on the other hand, do not allow accidental coreference, and neither do reduced pronouns, hence the computational task becomes much simpler, and children are more accurate. The fact that our lower proficiency L2ers were least accurate on full pronouns with referential antecedents suggests that the same computational burden arises in L2 acquisition, although not perhaps to the same degree, since our participants performed above chance on these sentences, unlike children.

To reiterate, the child language discoveries of a quantificational asymmetry and a clitic advantage found parallels in the performance of non-advanced L2ers. The fact that these same participants are at ceiling with reduced pronouns suggests that they know how to interpret such pronouns. The fact that they are highly accurate with quantified antecedents suggests that full pronouns are not always problematic. In other words, our lower proficiency learners do not have an underlying problem with all pronoun interpretation, but only with the difficult-to-compute cases, in consort with 6-year-old children acquiring English. No other theoretical account can explain the child—L2 learner parallel behavior.

In this respect, it is instructive to review Patterson et al.'s (2014) findings, in order to see whether their analysis can explain our results. These researchers attributed the performance of the L2ers in their experiments to a general preference for the non-local matrix subject to serve as the antecedent for a pronoun, even when this was not in fact the case for native speakers (as in the exceptional sentence types). However, such an explanation cannot account for our results. Our participants sometimes chose a local referential subject as antecedent and did so differentially in the case of full versus reduced pronouns.

As discussed above, Sorace and colleagues (as described in Sorace, 2011, 2016) have also proposed that certain problems relating to L2 pronoun interpretation (instability and overuse of overt subject pronouns in languages like Italian) may be attributed to differences in available processing resources, rather than differences in knowledge representation. The suggestion is that bilingual processing is less efficient than monolingual processing, either because of difficulties in accessing and integrating different kinds of linguistic knowledge or because of the availability of fewer cognitive resources in general. In our account of computational complexity, we follow Grodzinsky and Reinhart (1993) and Reinhart (2011) in assuming that, as far as Principle B is concerned, the complexity relates to the fact that speakers have to compute and compare two linguistic derivations and ultimately reject one of them, which sometimes proves difficult or impossible for language learners. In other words, our definition of computational complexity is somewhat narrower than Sorace's approach to availability or non-availability of certain processing resources. Nevertheless, we concur that an increased processing load is implicated in both cases; it is this processing load rather than representational difficulties that underlies the performance of our participants.

Coming back to our own findings, we must acknowledge two alternative explanations of the greater accuracy on reduced pronouns that we found. The first is that the L1s of our participants were French or Spanish, both languages with clitic pronouns, so participants could presumably have transferred the requisite knowledge that clitics do not allow accidental coreference from their native languages. In other words, their greater accuracy with reduced pronouns would reflect L1 transfer. On the other hand, if transfer is the main factor at work, it remains unexplained why participants had problems precisely in those areas where accidental coreference needs to be computed and rejected; given the L1s in this case, accidental coreference should not have been entertained at all and so no computational complexity should have arisen. In order to eliminate the possibility of transfer, a necessary next step will be to add participants whose L1 does not have clitic-like pronouns, in order to see whether they can recognize the clitic-like properties of English reduced pronouns, including the fact that the computational burden is decreased in such cases.

The second objection that might be raised to our study is that the English reduced pronoun *'m* can be ambiguous between *him* and *them*. Could it be that the participants interpret *'m* as *them*, then reject the sentence in stories like (8) for the wrong reason, accounting for their greater accuracy with reduced pronouns in the False scenarios? In fact, if this were the case, then one would expect inaccuracy (i.e., rejections) on the scenarios where the expected answer is True [see (10)], contrary to what was found. Clarification on this point could be provided by including an unambiguous reduced pronoun, such as *'r* (*her*) in subsequent studies. A related point is the possibility of participants not hearing the reduced pronoun at all, and treating the verbs as intransitive, e.g., *Harry sprayed* and *Every student drove to school*. In order to evaluate this possibility, we examined the eight verbs in our test. Only four of them could be used intransitively, suggesting that omission of the pronoun is not a likely explanation of our results. As pointed out above, the effects in our model take into account the possible by-item variation present in the data.

Another possible objection to our analysis here is that a computational burden would seem to imply a measurable processing cost but our experiment included only an untimed TVJT, a measure of interpretation, not processing. We concur with Sorace (2011: 20), who points out that it is a misconception to assume that processing cannot be addressed by means of offline tasks. The fact that lower proficiency participants in our study had a problem in interpreting ONLY those stimuli where a computational cost is implicated is already an indication of a processing cost. Furthermore, while children's difficulties with pronouns have primarily been documented with comprehension studies, a number of studies have confirmed that the same contrast holds in online processing as well. For example, Clackson et al. (2011) conducted a visual-world eye tracking study on the processing of both reflexives and pronouns by 6-to-9-year-old English-speaking children. The results suggest that both adults and children experienced competition and interference when they had to consider two same-gender antecedents for pronouns, one grammatically

permitted, namely the matrix subject, and one an inaccessible competitor antecedent, the embedded clause subject, in sentences such as *Peter was waiting outside the corner shop. He watched as Mr. Jones bought a huge box of popcorn for him over the counter.* However, adults were able to overcome this difficulty and provide accurate offline judgments, unlike children, whose judgments were significantly less accurate.

In SLA research, too, the recent eye tracking study of Kim et al. (2015) uncovered a sharp contrast in L2 learners' treatment of reflexives and pronouns (see Pronoun Interpretation in L2 Acquisition), partially consistent with our offline findings [since Kim et al. (2015) used only referential antecedents and full pronouns]. Thus both interpretation and processing findings point in the same direction: pronouns are more difficult to process than reflexives, although individuals with higher processing resources are capable of accomplishing the necessary reference set computation.

Although we look at offline pronoun interpretation by L2 learners and establish lower accuracy for full pronouns with referential antecedents, our approach predicts processing difficulties even when the learners make the right choice (rejecting local antecedents for pronouns). Such behavior is already previewed in results from Clackson et al.'s (2011) adult native speakers, who demonstrated difficulties reflected in online measures but managed to compensate in offline measures. The higher computational burden is predicted to be reflected in longer reaction times, or greater hesitation, even when participants succeed in reference set computation. We leave this prediction for further research.

## CONCLUSION

We have looked at how a proposed computational burden has effects on linguistic performance, such that L2 learners occasionally and temporarily make inaccurate judgments as to referents for pronouns, parallel to the difficulty reported for L1 acquirers. That this is not an issue of inappropriate representation is demonstrated by L2ers' accuracy with quantified antecedents and with reduced pronouns, in contrast to their performance on full pronouns with referential antecedents. Our findings take us beyond earlier L2 research on pronoun interpretation which has rarely looked at the quantificational asymmetry and never,

as far as we are aware, at the differential status of the pronoun. Our results support the claim that a computational burden is implicated in L2 as in L1, and that this burden can be overcome—advanced L2ers do not differ from native speakers in their ability to select the appropriate antecedents for pronouns, even when they have to compute and reject accidental coreference.

In keeping with the research topic “Language acquisition in diverse linguistic, cognitive and social circumstances,” we have uncovered a similar pattern of behavior between children acquiring their native language and L2ers at lower levels of proficiency, despite considerable diversity in acquisition circumstances (age, cognitive capacities, input, etc.). The child–adult parallels with respect to difficulties in engaging in reference set computation and eventual success in this domain are noteworthy. At the same time, there are child–adult differences: adult L2ers do not experience as severe a difficulty as children (around 83% accuracy compared to 53%). This is not surprising, given that adults presumably have computational abilities that are superior to those of children. What is of interest is that being an adult is not sufficient to remove the computational burden altogether.

## AUTHOR CONTRIBUTIONS

RS and LW contributed equally to the design of the experiment, the interpretation of the results, the discussion and the writing up of the article. NG performed the statistical analysis, wrote the results section and contributed to the interpretation discussion.

## FUNDING

This research was supported with grants from the Social Sciences and Humanities Research Council of Canada (#410-2011-0809) and the Fonds de recherche Québec: Société et culture (#2016-SE-188196).

## ACKNOWLEDGMENT

We thank Guilherme Garcia, Jeffrey Klassen, and Jiajia Su for their assistance.

## REFERENCES

- Adjémian, C. (1976). On the nature of interlanguage systems. *Lang. Learn.* 26, 297–320. doi: 10.1111/j.1467-1770.1976.tb00279.x
- Avrutin, S., and Thornton, R. (1994). Distributivity and binding in child grammar. *Linguist. Inq.* 25, 167–171.
- Avrutin, S., and Wexler, K. (1992). Development of principle B in Russian: coindexation at LF and coreference. *Lang. Acquis.* 2, 259–306. doi: 10.1207/s15327817la0204\_2
- Baauw, S. (2002). *Grammatical Features and the Acquisition of Reference: A Comparative Study of Dutch and Spanish*. New York, NY: Routledge.
- Baauw, S., and Cuetos, F. (2003). The interpretation of pronouns in Spanish language acquisition and breakdown: evidence for the “Delayed Principle B Effect” as a non-unitary phenomenon. *Lang. Acquis.* 11, 219–275. doi: 10.1207/s15327817la1104\_2
- Baauw, S., Escobar, M. A., and Philip, W. (1997). “A delay of principle B effect in Spanish speaking children: the role of lexical feature acquisition,” in *Proceedings of the GALA '97*, eds A. Sorace, C. B. Heycock, and R. Shillcock (Edinburgh: Human Communication Centre).
- Belletti, A., Bennati, E., and Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: evidence from near-native Italian. *Nat. Lang. Linguist. Theory* 25, 657–689. doi: 10.1007/s11049-007-9026-9
- Chien, Y.-C., and Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Lang. Acquis.* 1, 225–295. doi: 10.1207/s15327817la0103\_2
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris Publications.

- Clackson, K., Felser, C., and Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: evidence from eye-movements during listening. *J. Mem. Lang.* 65, 128–144. doi: 10.1016/j.jml.2011.04.007
- Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/S0142716406060024
- Conroy, A., Takahashi, E., Lidz, J., and Phillips, C. (2009). Equal treatment for all antecedents: how children succeed with principle B. *Linguist. Inq.* 40, 446–486. doi: 10.1162/ling.2009.40.3.446
- Corder, S. P. (1967). The significance of learners' errors. *Int. Rev. Appl. Linguist. Lang. Teach.* 5, 161–170. doi: 10.1515/iral.1967.5.1-4.161
- Crain, S., and McKee, C. (1985). "The acquisition of structural restrictions on anaphora," in *Proceedings of the NELS 15*, eds S. Berman, J. Choe, and J. McDonough (Amherst, MA: University of Massachusetts).
- Crain, S., and Thornton, R. (1998). *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, MA: MIT Press.
- Elbourne, P. (2005). On the acquisition of principle B. *Linguist. Inq.* 36, 333–365. doi: 10.1162/0024389054396908
- Evans, G. (1980). Pronouns. *Linguist. Inq.* 11, 337–362.
- Finer, D., and Broselow, E. (1986). "Second language acquisition of reflexive-binding," in *Proceedings of the NELS 16*, eds S. Berman, J.-W. Choe, and J. McDonough (Amherst, MA: University of Massachusetts), 154–168.
- Friedmann, N., Novogrodsky, R., and Balaban, N. (2010). "The effect of crossing dependencies on the acquisition of pronoun comprehension," in *Language Acquisition and Development: Generative Approaches to Language Acquisition*, eds A. Castro, J. Costa, M. Lobo, and F. Pratas (Newcastle, UK: Cambridge Scholars Press), 146–156.
- Gordon, P. (1996). "The truth-value judgment task," in *Methods for Assessing Children's Syntax*, eds D. McDaniel, C. McKee, and H. Cairns (Cambridge, MA: MIT Press).
- Grodzinsky, Y., and Reinhart, T. (1993). The innateness of binding and coreference. *Linguist. Inq.* 24, 69–101.
- Hartman, J., Sudo, Y., and Wexler, K. (2012). Principle B and phonologically reduced pronouns in child English. *Paper Presented at the 37th Annual Boston University Conference on Language Development (BUCLD 37)*, Boston, MA.
- Haznedar, B., and Schwartz, B. D. (1997). "Are there Optional Infinitives in child L2 acquisition?," in *Proceedings of the 21st Annual Boston University Conference on Language Development*, eds E. Hughes, M. Hughes, and A. Greenhill (Somerville, MA: Cascadia Press), 257–268.
- Heim, I. (1993). *Anaphora and Semantic Interpretation*. SFS-Report-07-93. University of Tübingen, Tübingen.
- Hirakawa, M. (1990). A study of the L2 acquisition of English reflexives. *Second Lang. Res.* 6, 60–85. doi: 10.1177/02676583900600103
- Jakubowicz, C. (1984). "On markedness and binding principles," in *Proceedings of the NELS 14*, eds C. Jones and P. Sells (Amherst, MA: University of Massachusetts), 154–182.
- Kayne, R. (1975). *French Syntax: The Transformational Cycle*. Cambridge, MA: MIT Press.
- Kim, E., Montrul, S., and Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: evidence from eye tracking. *Appl. Psycholinguist.* 36, 1317–1374. doi: 10.1017/S0142716414000307
- Koster, C. (1993). *Errors in Anaphora Acquisition*. Doctoral dissertation, Utrecht University, Utrecht.
- Lardiere, D. (2000). "Mapping features to forms in second language acquisition," in *Second Language Acquisition and Linguistic Theory*, ed. J. Archibald (Oxford: Blackwell), 102–129.
- Lee, D., and Schachter, J. (1997). Sensitive period effects in binding theory. *Lang. Acquis.* 6, 333–362. doi: 10.1207/s15327817la0604\_3
- Marinis, T., and Chondrogianni, V. (2011). Comprehension of reflexives and pronouns in sequential bilingual children: Do they pattern similarly to L1 children, L2 adults, or children with specific language impairment? *J. Neurolinguist.* 24, 202–212. doi: 10.1016/j.jneuroling.2010.02.009
- McKee, C. (1992). A comparison of pronouns and anaphors in Italian and English acquisition. *Lang. Acquis.* 2, 21–54. doi: 10.1207/s15327817la0201\_2
- Patterson, C., Trompelt, H., and Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Front. Psychol.* 5:147. doi: 10.3389/fpsyg.2014.00147
- Philip, W., and Coopmans, P. (1996). "The role of lexical feature acquisition in the development of pronominal anaphora," in *Amsterdam Series on Child Language Development*, Vol. 5, eds W. Philip and F. Wijnen (Amsterdam: Institute of General Linguistics at the University of Amsterdam).
- Prévost, P., and White, L. (2000). Missing surface inflection or impairment in second language? Evidence from Tense and Agreement. *Second Lang. Res.* 16, 103–133. doi: 10.1191/02676580067756046
- R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reinhart, T. (2006). *Interface Strategies: Optimal and Costly Computations*. Cambridge, MA: MIT Press.
- Reinhart, T. (2011). "Processing or pragmatics? Explaining the coreference delay," in *The Processing and Acquisition of Reference*, eds T. Gibson and N. Pearlmuter (Cambridge, MA: MIT Press), 157–194. doi: 10.7551/mitpress/9780262015127.003.0007
- Reinhart, T., and Reuland, E. (1993). Reflexivity. *Linguist. Inq.* 24, 657–720.
- Reuland, E. (2001). Primitives of binding. *Linguist. Inq.* 32, 439–492. doi: 10.1162/002438901750372522
- Reuland, E. (2011). *Anaphora and Language Design*. Cambridge, MA: MIT Press.
- Sag, I. (1976). *Deletion and Logical Form*. Doctoral dissertation, University of Maryland, College Park, MD.
- Selinker, L. (1972). Interlanguage. *Int. Rev. Appl. Linguist. Lang. Teach.* 10, 209–231. doi: 10.1515/iral.1972.10.1-4.209
- Sigurjónsdóttir, S. (1992). *Binding in Icelandic: Evidence from Language Acquisition*. Doctoral dissertation, University of California, Los Angeles, CA.
- Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguist. Approach. Biling.* 1, 1–33. doi: 10.1075/lab.1.1.01sor
- Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguist. Approach. Biling.* 6, 669–684. doi: 10.1075/lab.15055
- Sorace, A., and Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Lang. Res.* 22, 339–368. doi: 10.1007/s10936-015-9372-4
- Thomas, M. (1995). Acquisition of the Japanese reflexive *zibun* and movement of anaphors in Logical Form. *Second Lang. Res.* 11, 206–234. doi: 10.1177/026765839501100302
- Thornton, R., and Wexler, K. (1999). *Principle B, VP-ellipsis, and Interpretation in Child Grammar*. Cambridge, MA: MIT Press.
- Varlokosta, S. (2002). "(A)symmetries in the acquisition of principle B in typically-developing and specifically language impaired (SLI) children," in *The Process of Language Acquisition*, ed. I. Lasser (Berlin: Peter Lang Verlag Publishing), 81–98.
- White, L. (1998). Second language acquisition and Binding Principle B: child/adult differences. *Second Lang. Res.* 14, 425–439. doi: 10.1191/026765898675690231
- White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511815065
- Williams, E. S. (1977). Discourse and logical form. *Linguist. Inq.* 8, 101–139.
- Zesiger, P., Chillier-Zesiger, L., Arabatzi, M., Baranzini, L., Cronel-Ohayon, S., Franck, J., et al. (2010). The acquisition of pronouns by French children. A parallel study of production and comprehension. *Appl. Psycholinguist.* 31, 571–603. doi: 10.1017/S0142716410000147

**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 © 2017 Slabakova, White and Guzzo. 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.



# Production Is Only Half the Story — First Words in Two East African Languages

Katherine J. Alcock\*

Department of Psychology, Fylde College, Lancaster University, Lancaster, United Kingdom

## OPEN ACCESS

### Edited by:

Gary Morgan,  
City, University of London,  
United Kingdom

### Reviewed by:

Marilyn Vihman,  
University of York, United Kingdom  
Stanka A. Fitneva,  
Queen's University, Canada

### \*Correspondence:

Katherine J. Alcock  
k.j.alcock@lancaster.ac.uk

### Specialty section:

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

**Received:** 27 March 2017

**Accepted:** 12 October 2017

**Published:** 30 October 2017

### Citation:

Alcock KJ (2017) Production Is Only  
Half the Story — First Words in Two  
East African Languages.  
Front. Psychol. 8:1898.  
doi: 10.3389/fpsyg.2017.01898

Theories of early learning of nouns in children's vocabularies divide into those that emphasize input (language and non-linguistic aspects) and those that emphasize child conceptualisation. Most data though come from production alone, assuming that learning a word equals speaking it. Methodological issues can mean production and comprehension data within or across input languages are not comparable. Early vocabulary production and comprehension were examined in children hearing two Eastern Bantu languages whose grammatical features may encourage early verb knowledge. Parents of 208 infants aged 8–20 months were interviewed using Communicative Development Inventories that assess infants' first spoken and comprehended words. Raw totals, and proportions of chances to know a word, were compared to data from other languages. First spoken words were mainly nouns (75–95% were nouns versus less than 10% verbs) but first comprehended words included more verbs (15% were verbs) than spoken words did. The proportion of children's spoken words that were verbs increased with vocabulary size, but not the proportion of comprehended words. Significant differences were found between children's comprehension and production but not between languages. This may be for pragmatic reasons, rather than due to concepts with which children approach language learning, or directly due to the input language.

**Keywords:** language acquisition, vocabulary acquisition, Bantu languages, East Africa, Communicative Development Inventories

## INTRODUCTION

### What Are Children's First Spoken Words?

Children first learning to say words in a variety of languages tend to produce names for things (Gentner, 1982; Goldfield and Reznick, 1990; Au et al., 1994; Caselli et al., 1995; Bassano, 2000; Kauschke and Hofmeister, 2002; Bornstein et al., 2004; McDonough et al., 2011). Different schools of thought have put forward a variety of different explanations for this "noun bias."

Some authors suggest that this is due to children having a set of pre-existing biases including an object bias (Markman, 1990). Others conclude that biased output may be a consequence of input. Influences may include the types of referents, and their correspondences, that are present in child- (and adult-) directed speech (Gleitman et al., 2005). Both schools of thought appear to assume that there are robust and important differences between children's core knowledge of nouns and of other types of words.

Most data though are obtained from production rather than comprehension, so it is not certain that this is representative of children's underlying knowledge. In fact, even some classic papers including Goldin-Meadow et al. (1976) and Gentner (1982) suggest that bias toward nouns is possibly weaker in comprehension.

Given these theoretical suggestions, it is important to determine whether nouns and verbs are both represented in early word knowledge. Researchers need to investigate systematically a variety of languages, looking at both early comprehension and early production. It is also important to examine a variety of cultural settings. We cannot answer questions such as these by only carrying out research in Western settings or on European languages.

I will now assess further evidence for a predominance of nouns in spoken words; I will then turn to the first words comprehended. I will address differences and similarities between languages and cultures, as the literature so far has findings of both.

### Linguistic Variance in First Spoken Words

It is possible that noun bias is language-specific. Choi and Gopnik (1995) suggested that sentence-final verbs in Korean leads to verbs having greater salience. They conclude that Korean-learning language children learn verbs earlier than in other languages, in preference to nouns. Brown (1998) examined the spontaneous speech of children learning Tzeltal, a Mayan language, and concluded that an early appearance of verbs may be due to the richness of meaning carried by many verbs. She observes that at the one word stage children's utterances mainly consist of single verbs whose meanings are close to those conveyed by nouns in other languages. In addition, the very earliest words were observed, as in many languages, to be words for people.

Tardif and colleagues (Tardif et al., 1997, 1999) found that the proportion of nouns or verbs appearing in children's early vocabulary in English and Mandarin was dependent on the method used to collect data. They noted that more verbs appeared in spontaneous speech than in parent-completed vocabulary checklists. Tardif et al. nevertheless claim that Mandarin-learning children produce higher proportions of verbs than do English-learning children; they estimate that the Mandarin learners produce approximately equal numbers of verbs and nouns.

Xuan and Dollaghan (2013) also examined English and Mandarin, but in their case with bilingual children (hence reducing child effects) using Communicative Development Inventories (CDIs). More verbs were produced by the same child in early Mandarin than early English. This study only included children who already had 50 spoken words, a relatively high level of spoken vocabulary for this type of study.

Childers et al. (2007) also noted no excess of nouns in first words using parent-completed inventories in Ngas, a Chadic language spoken in Northern Nigeria. They found comparably high ratios of verbs in comprehension; parent-completed inventories are ideal for comparing production and comprehension.

### Linguistic Invariance in First Spoken Words

Some cross-linguistic data call these observed language specificities into question. The first group of studies quoted here have all used parent-completed inventories. Caselli et al. (1995) discuss the possibility that rich verb morphology, variable word order (including many verb-final child-directed utterances), and subject omission found in Italian might lead to earlier acquisition of verbs. Their data did not back this up: in both Italian and English, children used a preponderance of nouns in the first 50 words.

Looking only at the first 10 words produced, Tardif et al. (2008) suggest that names for people predominate in English, Mandarin, and Cantonese. Tardif et al. suggest that the classification of names for people as nouns is a mistake in this field. However, most other studies, cross-linguistic or otherwise, have concentrated on children with larger vocabularies.

Likewise, in Bornstein et al. (2004, an extensive cross-linguistic study of seven languages with differing sentence structures), a higher proportion of nouns than verbs was found in the vocabulary of 20 month old children, beyond the very first few words. The use of inventories may explain why this noun bias was found in all languages, even Korean. Other studies have also found no earlier verb production in Korean (Au et al., 1994; Kim et al., 2000) or Mandarin (Gentner, 1982).

Bornstein et al. (2004) suggest that child constraints (children's pre-existing assumptions or knowledge), common to every child learning language, may lead to the pattern of early learned nouns and later learned verbs. Caselli et al. (1995) also conclude from their comparative study that children learning different languages all respond in a characteristic way to nouns and verbs in the ambient language.

Finally, using spontaneous speech data, Stoll et al. (2012) found that in Chintang, a highly agglutinative language in which verb arguments are optional (so verbs appear more frequently than nouns), early language learners were still seen to produce a higher proportion of nouns to verbs than were adults. Stoll et al. (2012) suggest that the complex verb system in Chintang leads to a relative reduction in the number of verbs produced by children. This is in contrast to the argument by other authors of the reverse (Caselli et al., 1995; Childers et al., 2007). Stoll et al. (2012) also note that most studies, even those that analyse spontaneous speech samples including adult speech, do not assess noun:verb ratios in input.

### Cultural Variance in First Spoken Words

Children learning to speak the same language may not necessarily experience the same parenting or the same type of input. Bornstein and Cote (2005) examined vocabulary composition in 20-month-old children growing up in varied cultural locations: three languages (Spanish, Italian, and English) and two settings (urban and rural). Using the same methodology for all children and calculating nouns and verbs as a proportion of the words available for parents to select on the inventory, children aged 20 months studied by Bornstein and Cote (2005) in rural Italy produced equal proportions of nouns and verbs. This was in contrast to all other settings in this study, and also in contrast to the findings of the same researchers previously

(Bornstein et al., 2004). Bornstein and Cote (2005) suggest that there are differences in rural versus urban parents' use of verbs in child-directed speech – specifically more didactic use of verbs by rural parents (Camaioni et al., 1998).

In a more direct examination of cultural differences, Fernald and Morikawa (1993) found that American mothers' more frequent object labeling led to their infants having more nouns in their vocabularies than Japanese infants, whose mothers used more social routines. Tamis-LeMonda et al. (2014) note that across several cultures maternal responsiveness has been seen to vary in ways that are known to affect infants' acquisition of different types of language. Hence cultural factors influencing parenting affect both children's acquisition, and mode of use, of early words.

### Comprehension, Vocabulary Knowledge, and Pragmatic Constraints

I now turn to early vocabulary comprehension, which tends to have a greater proportion of verbs than does early vocabulary production (Bates, 1979). Caselli et al. (1995) suggest this may simply be an artifact of the types of experimental materials used to elicit comprehension behavior, which may include more verbs than the type of material used to elicit production.

Goldfield (2000), however, examined parents' elicitations of children's speech and actions in spontaneous speech samples. She found that there was a difference between action- and object-directed speech such that parents' elicitations of comprehension were more likely to be designed to lead to their child performing an action than indicating an object, while elicitations of child speech were directed toward production of a noun rather than a verb. This difference between action- and object-directed speech may help to explain the bias in children's early language production to nouns. This adds to the evidence that children's early word production is not wholly representative of their early word knowledge. Parents' utterance type seems to be very dependent on context: in book reading contexts, parents use more object-oriented utterances (Altınkamaş et al., 2014), whereas more action-oriented utterances are used in toy play, and this seems to occur even in very verb-oriented or very noun-oriented languages.

### Knowledge versus Comprehension

Many studies and reviews discuss 'learning' of first words without distinguishing between comprehension and production though most of the data on which these discussions rely are from studies of children's first spoken words. Many of the learning mechanisms proposed imply underlying 'knowledge' or 'learning' of lexical concepts (Markman, 1990; Gleitman et al., 2005). These discussions heavily imply that comprehension develops in direct parallel.

Examining data on early comprehension from parent-completed inventories shows that the proportion of verbs in early comprehended words is higher than the proportion in early-produced words (Bates, 1979; Caselli et al., 1995; Childers et al., 2007). In addition to ensuring that the data in the current study – on two languages which have been little studied to date – are compatible with those from previous studies, the current

study must ensure that the data from comprehension are directly comparable with those from production.

### Study Languages and Setting

Kiswahili and Kigirima are two Eastern Bantu languages, both spoken in rural coastal Kenya. The languages are very closely related and have extremely similar grammatical structure; both languages have the noun classes found in Bantu languages (similar to grammatical gender), with verbs, adjectives, possessives and other parts of speech agreeing with the noun class of nouns. The two languages have very similar verb morphology: the same grammatical features are marked on the verb in each language, with similar or identical verb affixes. Rich inflections, and especially rich verb inflections, are found in these languages as in others (such as Italian, Caselli et al., 1995).

However, the two languages are largely not mutually intelligible, despite a large number of cognates (possibly over 80%, Alcock et al., 2015). Census and informal estimates of the number of native speakers are around 15 million for Kiswahili and 900,000 for Kigirima (Simons and Fennig, 2017).

Like many other richly inflected languages, these languages have highly variable word order. The basic word order is SVO, any word order is grammatical though alternate word order is usually marked. Caselli et al. (1995) hypothesize that word order variation in the input may affect the timing of verbs acquisition in child vocabulary. Even where a language is in essence SVO, the verb is frequently in a salient sentence-initial or sentence final position.

Caselli et al. (1995) go on to hypothesize that subject omission may also lead to higher salience of verbs in infant-directed speech (IDS). Verbs will constitute a higher proportion of the input language for children. The two languages studied here both allow overt subject or object omission, increasing even further the proportion of utterances in IDS that consist only of a single verb.

Pragmatics may also influence children's vocabulary learning (Goldfield, 2000) when utterances in IDS expect an action or speech in response. Social expectations of even children speaking their first words, in this society, like other rural areas of developing countries, may include a high degree of obedience. This could mean that children hear more spoken commands, designed to result in child actions.

Some relevant data are available from other languages spoken in similar settings. Though these are not from the languages in question, it is possible to hypothesize from other data whether children are likely to hear commands and/or other types of speech in their input, and potentially to gain some idea of relative proportions of different types of input utterances.

Stoll et al. (2012) observed some prompts to repeat an utterance directed by adults toward children, in Chintang (rural Nepal). The Kenyan spontaneous speech samples also have some examples of older children eliciting repetitions (Alcock et al., 2012, 2015), and Rabain-Jamin (1998) observed this type of routine with toddlers in West Africa.

In both Rabain-Jamin's setting and another West African setting mothers differed from older children in the types of speech they directed to infants. While in both settings high proportions of imperatives or directives were used, mothers used

more declaratives while older children used more imperatives. Both mothers and older children described, and asserted, while older children used more Wh-questions (Nwokah, 1987; Rabain-Jamin, 1998). Rabain-Jamin also observed that mothers reported speech more often for younger toddlers (16–22 months) and prompted directly more with older toddlers (24–28 months).

Likewise in South Africa Kvalsvig et al. (1991) found that in Zulu- and Xhosa-speaking families, adults and older children used commands when speaking to pre-schoolers (age five), and pre-schoolers also used commands when talking to infants and toddlers. All interlocutors frequently used other types of speech acts, including informational and question acts. Roughly equal proportions of commands and information utterances were seen. Deen (2003) noted that around 30% of verbs in IDS Nairobi dialect Kiswahili were grammatical imperatives but did not quantify other utterance types.

Hence in similar cultures, commands – requests involving a verb and eliciting action – are heard in children's input and could potentially encourage verb comprehension in first words. Many other types of utterances are also heard, including direct prompts for repetition.

## Predictions

Taking into account findings from a language with similarly complex and salient verbs (Caselli et al., 1995), and data from this setting and similar societies where commands are given at least as often as in European IDS (Goldfield, 2000), I hypothesize that children learning these two closely related Eastern Bantu languages will produce more nouns in their first spoken words than other categories. In contrast the noun bias predicted for production will be smaller or absent in the case of comprehension.

I hypothesize that this bias in production is due at least in part to factors, possibly input factors, that differentially affect spoken words – in other words the bias is not present in underlying early word knowledge itself. Although the study design does not allow direct assessment of mechanisms that could determine the source of any difference between production and comprehension, a smaller or non-existent noun bias in comprehension will necessarily imply the same in knowledge.

## Methodology

The method used needs to work in this setting and be comparable both across production and comprehension and with previous studies. Parent-completed inventories have been validated both for comprehension and production of vocabulary (Fenson et al., 1994). In particular, parents can use them to accurately report comprehension vocabulary (Mills et al., 1993, 1997; Schafer, 2005; Styles and Plunkett, 2009).

Using parent-completed inventories to collect cross-linguistic vocabulary data, Bornstein et al. (2004) examined production vocabulary only, while Caselli et al. (1995) looked at production and comprehension. In order to ensure cross-linguistic comparability, the current study will rely for the most part on a replication of the methods of Caselli et al. (1995; see below for more details), but adding analyses using the Bornstein et al. (2004) method of correcting for the number of opportunities

that parents have to choose a word in any given category. Since there are more nouns than verbs in most parent-completed vocabulary inventories, the number of nouns a child knows may be artificially inflated if this correction is not carried out.

Both Caselli et al. (1995) and Bornstein et al. (2004) administered CDIs in written format. The Kenyan CDIs were necessarily administered in interview format.

## MATERIALS AND METHODS

### Participants and Materials

A total of 208 families with children aged 8–20 months (mean 12.99 months, SD 2.91), resident in Kilifi District, Coast Province, Kenya, took part in the study. Families were recruited through a periodic census of villages and homesteads in the area. Of these families 63 were predominantly Kiswahili-speaking and 145 were predominantly Kigirima-speaking. Speakers of the two languages are usually resident in different villages and follow different religions, so children are exposed primarily to their home language in their village and at social occasions. Where more than one of the languages, or another language, was spoken by adults to children, these children were excluded from the study. However, most adults in the study area speak at least a little of both languages plus some English, so some code-switching occurs in these primarily monolingual homes.

Families were interviewed verbally with the Kiswahili or Kigirima version of the MacArthur-Bates Communicative Development Inventory – Words and Gestures (Fenson et al., 1994), constructed and validated for this community (Alcock et al., 2015). Assessment of both production and comprehension with the CDI were found to have external validity. Validation included comparison of parent report of comprehension with children's communicative behavior (gesture and object name comprehension) in a session at children's homes. Note in particular we found a relationship between parent report of comprehension of specific words on the CDI and children's comprehension in a testing session of those particular items (significant at the one-tailed level with  $N = 17$ ). We also validated vocabulary production in older toddlers against spontaneous speech production taken from home recordings, and against a picture vocabulary test. This gives confidence that the tool is valid for measurement both of comprehension and production.

An interview version of the CDI has also been validated against the Bayley Scales of Infant Development (Mental) in another, similar illiterate population (Hamadani et al., 2010). Data for the current study were collected as part of larger study investigating the effect of HIV exposure on infant development; the data presented here are from children who were not known to be exposed to HIV.

### Vocabulary Categories and Word Ranking

The number of words in each vocabulary category on the inventory is shown in **Table 1**. The inventory has a total of 292 vocabulary items. These were categorized using the method of

**TABLE 1 |** Number of words in each lexical category on the vocabulary inventory.

|                    | Sound effects       | Animals                                                                                      | Vehicles                    | Toys                 | Foods              | Clothes              | Body parts          | Small household objects                                                                          |
|--------------------|---------------------|----------------------------------------------------------------------------------------------|-----------------------------|----------------------|--------------------|----------------------|---------------------|--------------------------------------------------------------------------------------------------|
| Number of words    | 15                  | 15                                                                                           | 5                           | 10                   | 39                 | 13                   | 15                  | 34                                                                                               |
|                    | Furniture and rooms | Outdoor items                                                                                | Places to go                | People's names       | Games and routines | Verbs                | Adjectives          | Function words                                                                                   |
| Number of words    | 11                  | 18                                                                                           | 10                          | 14                   | 12                 | 56                   | 15                  | 10                                                                                               |
| Broader categories | Category: Nominals  | Sub-category: Common nouns (includes some items e.g., <i>shop</i> from <i>Places to Go</i> ) | Sub-category: Sound effects | Sub-category: People | Category: Routines | Category: Predicates | Sub-category: Verbs | Sub-category: Adjectives                                                                         |
|                    | 194                 | 165                                                                                          | 15                          | 14                   | 12                 | 71                   | 56                  | 15                                                                                               |
|                    |                     |                                                                                              |                             |                      |                    |                      |                     | Category: Function words (includes remaining items e.g., <i>there</i> from <i>Places to Go</i> ) |
|                    |                     |                                                                                              |                             |                      |                    |                      |                     | 15                                                                                               |

Caselli et al. (1995). This method initially categorizes words into four broader categories, followed by seven narrower categories: Nominals (Common nouns, Proper nouns, and Sound effects), Routines, Predicates (Verbs and Adjectives), and Function words.

For each language and for production and comprehension the most frequent 50 words produced and the most frequent 50 words comprehended (the first 50 words by rank) were noted. This replicates the methods of Caselli et al. (1995).

## RESULTS

### Categorization of First Fifty Words

**Table 2** shows the categorisation of all words ranked under 50 in production and **Table 3** shows the same figures for comprehension, by language. Exactly 50 words were ranked between 1 and 50 in comprehension for both languages. However, because several words can be (and were) ranked equally, the number of words ranked under 50 for production is not the same in each language. This means that the number of words with this rank is greater than 50 (63 for Kigirima and 57 for Kiswahili). The total numbers in each word category in production are hence shown in **Table 2** scaled down to 50. The vocabulary items ranked 1 through 50 in each language in comprehension, 1 through 46 in Kigirima and 1 through 44 in Kiswahili in production, are shown in the Appendix, together with a translation equivalent.

Chi-square analysis revealed no significant differences in the categorisation of first words between the two languages, either in comprehension or in production, and with either broader or narrower categories. In addition, *t*-tests showed no differences in the total number of words produced or comprehended by children learning the two different languages; for production vocabulary  $t(206) = 0.751$  and for comprehension  $t(206) = 0.873$ . Given the extremely high rate of cognates and the very closely related nature of the two languages, further data shown are from both languages, combined. It can be seen from these tables that, as in English and Italian, the majority of the earliest 50 words produced by children are nominals.

### Quantitative Vocabulary Growth in Comprehension and Production with Age

Data from only small numbers of children over the age of 16 months (the target maximum age for typically developing children for the original Words and Gestures inventory) were available, so for age analyses such children are excluded from the dataset. Mean vocabulary size of older children was within the range for the younger children, so their data were included in analyses of vocabulary categories by vocabulary size.

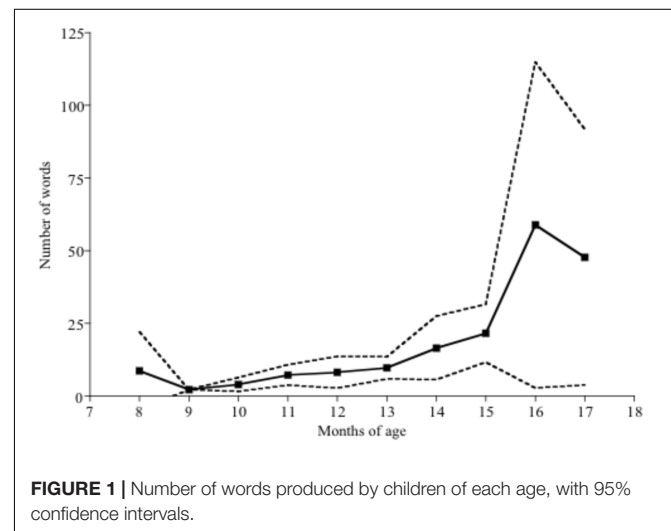
The mean number of words produced and comprehended by children of each month of age can be seen in **Figures 1, 2** respectively. Both production and comprehension vocabulary correlated significantly with children's ages in months. For production vocabulary  $r(184) = 0.33$ ,  $p < 0.001$  and for comprehension vocabulary  $r(184) = 0.50$ ,  $p < 0.001$ . Further details of the relationship between age and vocabulary are discussed in Alcock et al. (2015).

**TABLE 2 |** Highest ranked 50 words in each language, categorized by word class – Count of words for Production.

| Category – broader categories | Category – narrower categories | Language |           | Scaled to 50 words |           |
|-------------------------------|--------------------------------|----------|-----------|--------------------|-----------|
|                               |                                | Kigirama | Kiswahili | Kigirama           | Kiswahili |
| Function words                |                                | 1        | 0         | 1                  | 0         |
| Nominals                      |                                | 54       | 49        | 43                 | 43        |
|                               | Common nouns                   | 33       | 31        | 26                 | 27        |
|                               | People                         | 11       | 10        | 9                  | 9         |
|                               | Sound effects                  | 10       | 8         | 8                  | 7         |
| Predicates                    |                                | 2        | 3         | 2                  | 3         |
|                               | Adjectives                     | 1        | 2         | 1                  | 2         |
|                               | Verbs                          | 1        | 1         | 1                  | 1         |
| Routines                      |                                | 6        | 5         | 5                  | 4         |

**TABLE 3 |** Highest ranked 50 words in each language, categorized by word class – Count of words for Comprehension.

| Category – broader categories | Category – narrower categories | Language |           |
|-------------------------------|--------------------------------|----------|-----------|
|                               |                                | Kigirama | Kiswahili |
| Function words                |                                | 0        | 1         |
| Nominals                      |                                | 35       | 32        |
|                               | Common nouns                   | 24       | 23        |
|                               | People                         | 5        | 4         |
|                               | Sound effects                  | 6        | 5         |
| Predicates                    |                                | 14       | 14        |
|                               | Adjectives                     | 1        | 2         |
|                               | Verbs                          | 13       | 12        |
| Routines                      |                                | 1        | 3         |



## Change in Categories as Vocabulary Grows

Analyses of the relationship between vocabulary categories and vocabulary size were planned and carried out as follows:

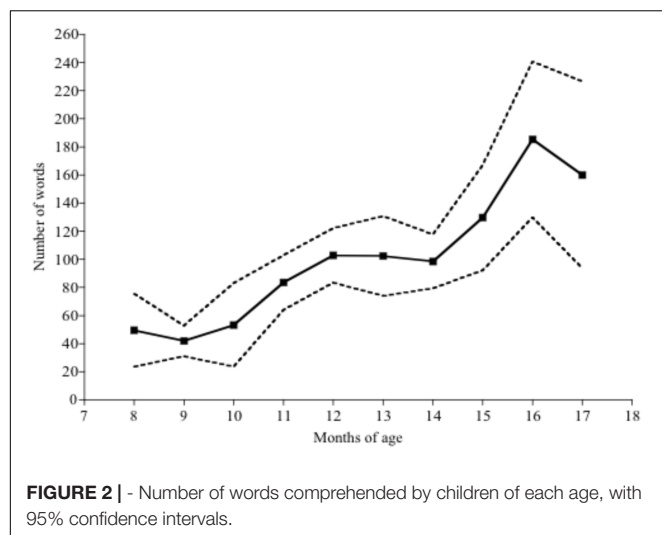
- (1) Separate analyses of comprehension and production vocabularies:
- (a) Simple percentage of all words known in comprehension and production in *broadier* (Caselli et al., 1995's Nominals, Predicates, Routines and Function words) and
- (b) *narrower* (Common Nouns, Verbs, Adjectives and Function Words in Caselli et al. (1995) - Closed Class in Bornstein et al., 2004) word categories.
- (c) Analyses computed as a proportion of the words in each category in the checklist – “chances to choose each category of word” – as in Bornstein et al. (2004) – *broadier* categories of words, comprehension and production, for children with different vocabulary sizes and
- (d) *narrower* categories of words, comprehension and production, for children with different vocabulary sizes.

- (2) Combined analyses of comprehension and production vocabulary as in (c) and (d) only (proportion of chances – Bornstein et al., 2004).

## Analysis 1 – Nouns and Verbs in Early Production and Comprehension Vocabularies

Analysis 1a and 1b. Analysis of raw proportions of verbs and nouns in children's comprehension and production vocabularies. Production: Here, nouns predominate in early production vocabulary with verbs forming a much smaller proportion of children's early words – less than 10% in all vocabulary levels up to the 50 word production point. Data on both broader and narrower categories of production vocabulary can be seen in **Table 4**. As can be seen from this Table, the proportion of nominals in production starts at 96% in the smallest vocabulary group (1–5 words) and drops to 75% in the largest group (51+ words).

*Comprehension* can be seen in **Table 5**. Nominals are a smaller proportion of the vocabulary at all vocabulary sizes and verbs are over 15% of the vocabulary even at the smallest vocabulary size.



For the earliest stages (before the Kenyan children reach 50 words) the proportion of words that are verbs is very low in production. However after this stage (after 50 words) Eastern Bantu-learning children start to produce a mean of 11.8% of their words as verbs.

In comprehension, at a vocabulary level of under 21 words, the proportion of words that Eastern Bantu-learning children understood were 16.2% verbs. Nevertheless, in both production and comprehension the majority of words are nominals, at all vocabulary levels.

## Categorisation for Analyses 1c and 1d: Vocabulary Size Category Assignment

Bornstein et al. (2004) analyzed children's production vocabulary by calculating vocabulary in each category of words as a proportion of the number of chances parents have to choose a word of that category – since in early vocabulary inventories, there are more nouns than other word types to choose from. The categories in the current study correspond to Bornstein et al.'s (2004) Nouns, Verbs, Adjectives and Closed Class.

Bornstein et al. (2004) analyzed data from somewhat older children (20 months) than in this paper, with larger vocabulary

sizes. **Table 6** therefore shows vocabulary in each category as a proportion of available chances for children in the production vocabulary groups used in analyses 1a and 1b (ranging from 1 to 5 words to 51+ words). **Table 7** shows comprehension. Note these are not the same groups as in Bornstein et al. (2004) due to the smaller vocabulary size of the Kenyan children.

Children with 1–50 spoken words produced a mean of 4% of the nominals on the inventory compared with a mean of less than 1% of the verbs. Likewise children with larger vocabularies produced 44% of the nominals and only 24% of the verbs.

In comprehension, children with both smaller (<200 words) and larger (>200 words) comprehension vocabularies were reported to comprehend almost equal proportions of the nominals on the inventory (31% for the smaller vocabulary group and 87% for the larger vocabulary group) and verbs (34 and 87% respectively).

**Analysis 1c.** Analysis of broader categories of words produced and comprehended as a percentage of chances to choose those words, for children with different vocabulary sizes. ANOVAs were carried out to compare proportions of words on the inventory in each category produced versus comprehended by children in different vocabulary groups. These used the original four broad categories Nominals, Predicates, Routines and Function words.

For *production*, a significant main effect was found of word category,  $F(3,184) = 33.18$ ,  $p < 0.001$ ,  $\eta^2 = 0.15$ . A significant interaction between word category and vocabulary group,  $F(12,184) = 5.00$ ,  $p < 0.001$ ,  $\eta^2 = 0.10$  was also found. For all ANOVAs post-hoc pairwise comparisons with Bonferroni corrections were carried out.

These pairwise comparisons showed that although the proportion of nominals did not differ significantly from that of routines, all other pairs were significantly different: parents reported a significantly higher proportion of nominals than predicates or function words, of routines than of predicates and function words, and of function words than of predicates, were produced. The differences between word categories became smaller as vocabularies became bigger, however.

For *comprehension*, a significant main effect of word category was again seen,  $F(3,202) = 9.85$ ,  $p < 0.001$ ,  $\eta^2 = 0.05$ , but no significant interaction between vocabulary size and word category. As with the raw data analysis above,

**TABLE 4 |** Percentages of vocabulary consisting of words in each category, across vocabulary sizes – Production.

|                  | Number of words in production vocabulary (N) |           |            |            |          | Total |
|------------------|----------------------------------------------|-----------|------------|------------|----------|-------|
|                  | 1–5 (82)                                     | 6–10 (35) | 11–20 (41) | 21–50 (16) | 51+ (15) |       |
| % Nominals       | 96.0                                         | 91.0      | 88.1       | 84.5       | 75.0     | 90.8  |
| % Common nouns   | 2.7                                          | 13.6      | 23.6       | 43.4       | 59.1     | 17.0  |
| % People         | 78.0                                         | 40.2      | 30.4       | 21.2       | 8.7      | 50.7  |
| % Sound effects  | 15.5                                         | 37.1      | 34.2       | 20.0       | 7.2      | 23.2  |
| % Routines       | 3.8                                          | 8.3       | 9.5        | 6.8        | 5.5      | 6.2   |
| % Predicates     | 0.0                                          | 0.5       | 1.8        | 4.5        | 15.8     | 2.1   |
| % Verbs          | 0.0                                          | 0.0       | 1.4        | 2.5        | 11.8     | 1.4   |
| % Adjectives     | 0.0                                          | 0.5       | 0.4        | 2.0        | 4.0      | 0.7   |
| % Function words | 0.3                                          | 0.3       | 0.6        | 4.2        | 3.7      | 0.9   |

**TABLE 5 |** Percentages of vocabulary consisting of words in each category, across vocabulary sizes – Comprehension.

|                  | Number of words in comprehension vocabulary (N) |            |             |              |              |           | Total |
|------------------|-------------------------------------------------|------------|-------------|--------------|--------------|-----------|-------|
|                  | 0–20 (20)                                       | 21–50 (43) | 51–100 (62) | 101–150 (40) | 151–200 (26) | 201+ (17) |       |
| % Nominals       | 78.6                                            | 68.9       | 64.6        | 65.8         | 67.5         | 67.0      | 67.6  |
| % Common nouns   | 38.7                                            | 46.0       | 48.3        | 53.9         | 57.6         | 58.2      | 50.0  |
| % People         | 21.1                                            | 12.6       | 7.7         | 6.3          | 5.3          | 4.8       | 9.2   |
| % Sound effects  | 18.8                                            | 10.6       | 8.6         | 5.6          | 4.6          | 4.1       | 8.5   |
| % Routines       | 3.4                                             | 4.8        | 4.6         | 4.0          | 3.6          | 3.7       | 4.2   |
| % Predicates     | 18.1                                            | 24.1       | 28.4        | 27.3         | 25.9         | 25.9      | 25.8  |
| % Verbs          | 16.3                                            | 20.2       | 24.4        | 22.3         | 21.7         | 20.8      | 21.7  |
| % Adjectives     | 1.8                                             | 3.9        | 4.0         | 5.0          | 4.4          | 5.1       | 4.1   |
| % Function words | 0.0                                             | 2.2        | 2.4         | 2.9          | 2.9          | 3.3       | 2.4   |

**TABLE 6 |** Categories of words in production vocabulary at different vocabulary sizes expressed as a proportion of chances to choose each category.

| Proportion of words on inventory | Number of words in production vocabulary (N) |           |            |            |            |          | Total |
|----------------------------------|----------------------------------------------|-----------|------------|------------|------------|----------|-------|
|                                  | 1–5 (82)                                     | 6–10 (35) | 11–20 (41) | 21–50 (16) | 1–50 (174) | 51+ (15) |       |
| Nominals                         | 0.01                                         | 0.04      | 0.07       | 0.14       | 0.04       | 0.44     | 0.07  |
| Common nouns                     | 0.00                                         | 0.01      | 0.02       | 0.08       | 0.01       | 0.41     | 0.05  |
| People                           | 0.12                                         | 0.21      | 0.29       | 0.42       | 0.21       | 0.60     | 0.24  |
| Sound effects                    | 0.03                                         | 0.20      | 0.31       | 0.38       | 0.16       | 0.47     | 0.19  |
| Routines                         | 0.01                                         | 0.05      | 0.10       | 0.17       | 0.06       | 0.44     | 0.09  |
| Predicates                       | 0.00                                         | 0.00      | 0.00       | 0.02       | 0.00       | 0.25     | 0.02  |
| Verbs                            | 0.00                                         | 0.00      | 0.00       | 0.01       | 0.00       | 0.24     | 0.02  |
| Adjectives                       | 0.00                                         | 0.00      | 0.00       | 0.04       | 0.01       | 0.31     | 0.03  |
| Function words                   | 0.00                                         | 0.00      | 0.01       | 0.13       | 0.01       | 0.38     | 0.04  |

**TABLE 7 |** Categories of words in comprehension vocabulary at different vocabulary sizes expressed as a proportion of chances to choose each category.

|                | Number of words in comprehension vocabulary (N) |            |             |              |              |             | Total |
|----------------|-------------------------------------------------|------------|-------------|--------------|--------------|-------------|-------|
|                | 0–20 (20)                                       | 21–50 (43) | 51–100 (62) | 101–150 (40) | 151–200 (26) | 1–200 (191) |       |
| Nominals       | 0.05                                            | 0.15       | 0.26        | 0.45         | 0.66         | 0.31        | 0.35  |
| Common nouns   | 0.03                                            | 0.12       | 0.23        | 0.43         | 0.66         | 0.28        | 0.33  |
| People         | 0.16                                            | 0.34       | 0.40        | 0.54         | 0.66         | 0.43        | 0.46  |
| Sound effects  | 0.13                                            | 0.28       | 0.41        | 0.45         | 0.54         | 0.38        | 0.40  |
| Routines       | 0.03                                            | 0.16       | 0.27        | 0.41         | 0.54         | 0.28        | 0.32  |
| Predicates     | 0.04                                            | 0.13       | 0.29        | 0.47         | 0.65         | 0.32        | 0.36  |
| Verbs          | 0.04                                            | 0.14       | 0.32        | 0.49         | 0.69         | 0.34        | 0.38  |
| Adjectives     | 0.02                                            | 0.10       | 0.19        | 0.41         | 0.52         | 0.24        | 0.29  |
| Function words | 0.00                                            | 0.09       | 0.17        | 0.35         | 0.52         | 0.22        | 0.27  |

for comprehension there are no differences in vocabulary composition as comprehension vocabulary increases. Here pairwise comparisons showed significantly higher proportions of nominals than routines or function words, and of predicates than routines or function words, were comprehended, but there was no significant difference between the proportion of nominals and of predicates that were comprehended, nor between routines and function words.

Analysis 1d. Analysis of narrower categories of words produced and comprehended as a percentage of chances to choose those words, for children with different vocabulary sizes. Analyzed in this way, with narrower groups of words

comparable to the analyses of Bornstein et al. (2004) the data also show a predominance of nouns in first words, especially in production. The sample in the current dataset is biased toward children with smaller vocabularies, so the proportions are not completely comparable to those of Bornstein et al. (2004) Nevertheless, taking the children with 51 or more words as the median group in the Bornstein et al. (2004) 'smaller vocabularies' group, the figure of slightly less than twice as many nouns (compared to noun-opportunities) versus verbs (compared to verb-opportunities), is similar to the figures for most of the languages in the Bornstein et al. (2004) data.

The picture for comprehension is different, however – children with smaller comprehension vocabularies – 1–200 words – comprehended 28% of the possible common nouns and a higher proportion, 34%, of the possible verbs. Children with comprehension vocabularies over 200 comprehended 89% of the possible common nouns and 87% of the possible verbs, though in this group a ceiling effect may be operating.

ANOVAs were carried out to examine growth of vocabulary in these categories as overall vocabulary sizes grow. For production, a significant effect of category,  $F(3,183) = 14.53$ ,  $p < 0.001$ ,  $\eta^2 = 0.07$  and an interaction between category and vocabulary level,  $F(12,183) = 4.31$ ,  $p < 0.001$ ,  $\eta^2 = 0.09$  were found. As children's vocabularies grew, the proportions of different word classes produced became more similar.

For comprehension, again a significant effect of word category,  $F(3,201) = 12.60$ ,  $p < 0.001$ ,  $\eta^2 = 0.06$  was found, but as above there was no interaction; there is no change in the proportions of words in different categories as vocabulary grows. Data from these comparisons for production and comprehension can be seen in **Figures 3, 4** respectively.

## Analysis 2 – Combined analysis of Comprehension and Production

Grand ANOVAs (combining previous analyses) were carried out to compare the proportions of the words on the checklist that can be seen in children's production versus comprehension at different vocabulary sizes. As only one measure of vocabulary size can be used for this analysis, comprehension vocabulary size was chosen – all children had a comprehension vocabulary of 1 or more, while many had a production vocabulary of zero, reducing the variance. This means that these analyses are not precisely comparable to the separate analyses above.

Analysis 2c. Broader categories - Nominals, Predicates, Routines and Function words – Comparison of production and comprehension. Significant main effects of modality, word class, and vocabulary group were found, as well as significant interactions between modality and both vocabulary group and word class, and a three-way interaction between modality, vocabulary group and word class. Results of these two grand ANOVAs are shown in **Table 8**.

Planned comparisons showed differences between Nominals and Predicates (mean difference = 0.03,  $SE = 0.005$ ,  $p < 0.001$ ), and between Nominals and Function Words (mean difference = 0.06,  $SE = 0.016$ ,  $p = 0.002$ ).

Analysis 2d. Narrower categories - Common Nouns, Verbs, Adjectives, and Function words – Comparison of production and comprehension. Main effects were found of modality (comprehension versus production), word class and vocabulary group in addition to interactions between modality and both vocabulary group and word class. Planned comparisons showed significant differences between Common Nouns and Adjectives (mean difference = 0.04,  $SE = 0.007$ ,  $p < 0.001$ ) and between Verbs and Adjectives (mean difference = 0.04,  $SE = 0.007$ ,  $p < 0.001$ ).

Hence when comprehension and production are compared directly, the above findings are confirmed. As children's

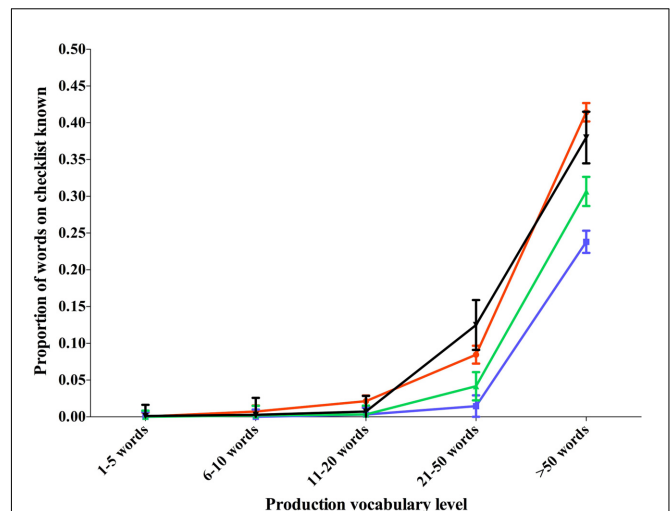


**FIGURE 3 |** Proportion of words in different categories produced by children of different vocabulary levels.

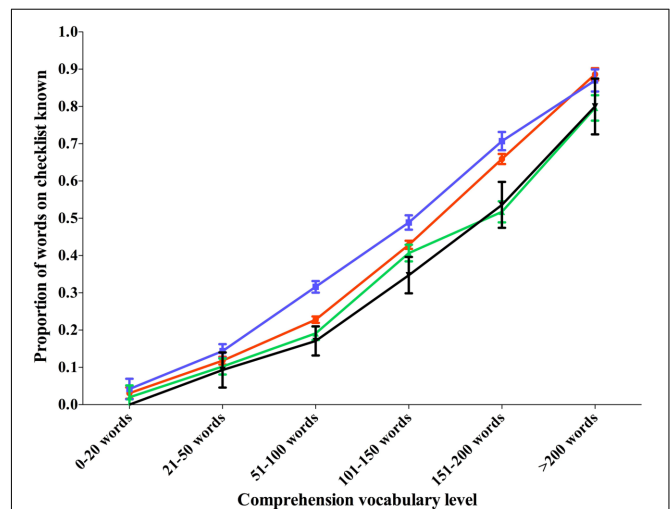


**FIGURE 4 |** Proportion of words in different categories comprehended by children of different vocabulary levels.

vocabulary gets bigger, the proportion of words that they produce in different classes changes, but the proportion of words that they comprehend in different classes does not.

## DISCUSSION

### The First Words that Children Say

When comparable techniques are used to investigate children whose input language varies, the first words that children say are predominantly nouns. This has been found in children who hear a variety of European, Asian and now African languages. The two extremely closely related Eastern Bantu languages studied here both allow sentences that consist of a single, highly inflected verb, as do Spanish or Italian. Such single-verb sentences may

**TABLE 8 |** Analyses of variance examining proportion of words known in different word classes in both comprehension and production modalities, as a function of vocabulary size.

| Source                                     | d.f | F       | $\eta^2$        | P      |
|--------------------------------------------|-----|---------|-----------------|--------|
| <b>Between subjects</b>                    |     |         |                 |        |
| Vocabulary group                           | 5   | 129.53  | 0.76            | <0.001 |
|                                            |     | 150.48  | 0.79            | <0.001 |
| <b>Within subjects</b>                     |     |         |                 |        |
| Modality                                   | 1   | 983.24  | 0.83            | <0.001 |
|                                            |     | 1104.91 | 0.85            | <0.001 |
| Word class                                 | 3   | 8.16    | 0.04            | <0.001 |
| Vocabulary group by modality               | 5   | 69.74   | 0.63            | <0.001 |
|                                            |     | 90.73   | 0.69            | <0.001 |
| Vocabulary group by word class             | 15  |         | Not significant |        |
| Modality by word class                     | 3   | 18.68   | 0.08            | <0.001 |
|                                            |     | 20.21   | 0.09            | <0.001 |
| Vocabulary group by modality by word class | 15  | 2.61    | 0.06            | 0.001  |
|                                            |     |         | Not significant |        |

Cells show ANOVA 1 (Common Nouns, Routines, Predicates, Function Words) in the upper half and ANOVA 2 (Nominals, Verbs, Adjectives and Function Words) in the lower half. Degrees of freedom are the same for both ANOVAs.

be even more common in Bantu than in Romance languages, since in Bantu languages subjects and objects can be represented as verb affixes. However, even single-verb sentences and highly variable word order do not lead children to produce verbs in large proportions in their first spoken words. Likewise, documented elicitations of other types of words from infants by older children might have led to lower proportions of nouns in first spoken words, but this is not the case.

This predominance of nouns in first spoken words holds up for children with vocabularies from 1 to 5 words up to more than 50 words. Early vocabulary checklists tend to contain a large predominance of nominals but nouns also predominate when the number of words in each category was analyzed as a proportion of chances to choose those categories of words in those categories. The results are the same, however, the words are categorized, too, whether as nouns versus verbs, adjectives and function words or whether of nominals versus predicates/function words.

As children's spoken vocabularies grow, the proportion of words in different categories do change, however: there is a significant interaction between spoken vocabulary size and the proportion of words in each vocabulary categories. It is necessary to be cautious, though, in definitely categorizing children's first spoken words as verbs or nouns. Even in languages where the surface forms of these are different, children may use a surface noun to represent an action, or a surface verb to represent an object associated with an action.

## The First Words that Children Understand

The picture is very different in comprehension, however. In the earliest words comprehended (1–20 words) nominals are also very common, but a higher percentage of words comprehended than words produced are verbs. At larger comprehension vocabularies, the proportion of words comprehended that are

verbs increases slightly. Likewise, when analyzing percentage of chances to choose words in different categories, children at these levels of comprehension understand almost exactly the same percentage of the nouns and verbs on the checklist. As comprehension increases there is no significant change in the proportions of different types of words: the relative proportions of words in different classes remains the same as vocabulary grows.

## Directly Comparable Studies from Other Languages

Although the differences seen here between nouns and verbs and between production and comprehension are very similar to the differences found by Caselli et al. (1995) in both US English and Italian, production data from these Bantu languages may be more similar to the data from Italian than to that from US English. For example, among the first 50 words spoken in Italian are 8 words for people, compared with 9 in Kigirama/Kiswahili and just 4 in US English. As suggested by Caselli et al. (1995), it is reasonable to conclude that this reflects the frequent contact which children in some societies have with extended family members.

There is a hint that verbs may be growing faster in early Kenyan children's production vocabularies than in either US or Italian children's production vocabularies. Children whose spoken vocabularies are greater than 50 words say fewer verbs in either US English or Italian than children learning Kiswahili or Kigirama. When the number of words in each category was taken into account, Kenyan children in this spoken vocabulary group produced 41% of the common nouns on checklists and 24% of the verbs (a ratio of approximately 1.7:1). Looking at the Bornstein et al. (2004) data from older infants, for those with spoken vocabularies in the 51–100 word range, the ratio of noun:verb as a proportion of chances to choose words is very similar.

**TABLE 9 |** Cross-linguistic comparisons of noun and verb use by children in the smallest and largest vocabulary groups.

| Language                          | Comprehension – percentage of child's vocabulary. Child comprehension 20 words or less |       | Production – percentage of child's vocabulary. Child production 1–5 words |       |
|-----------------------------------|----------------------------------------------------------------------------------------|-------|---------------------------------------------------------------------------|-------|
|                                   | Nouns                                                                                  | Verbs | Nouns                                                                     | Verbs |
| <i>Smallest vocabulary groups</i> |                                                                                        |       |                                                                           |       |
| Kiswahili/Kigiriana               | 78.6                                                                                   | 16.2  | 96.0                                                                      | 0.0   |
| Italian                           | 66.8                                                                                   | 6.9   | 80.4                                                                      | 1.3   |
| English                           | 60.4                                                                                   | 6.8   | 91.0                                                                      | 0.5   |
| Language                          | Comprehension – % of child's vocabulary. Child comprehension more than 200 words       |       | Production – % of child's vocabulary. Child production more than 50 words |       |
|                                   | Nouns                                                                                  | Verbs | Nouns                                                                     | Verbs |
| <i>Largest vocabulary groups</i>  |                                                                                        |       |                                                                           |       |
| Kiswahili/Kigiriana               | 67.0                                                                                   | 20.8  | 75.0                                                                      | 11.8  |
| Italian                           | 60.7                                                                                   | 17.5  | 72.6                                                                      | 6.8   |
| English                           | 61.8                                                                                   | 16.0  | 73.6                                                                      | 4.5   |

Data are from (Caselli et al., 1995).

Between-language comparisons of the proportion of children's vocabulary that is in each category are shown in **Table 9**. As discussed above, the proportion of nouns to verbs in early comprehension vocabulary does not seem to change as children increase their vocabularies in the Kenyan languages.

Caselli et al. (1995) suggest that the excess of nouns over verbs in the construction of CDIs represents both an accurate reflection of the composition of adult vocabulary and of children's early vocabulary – that children indeed first learn more nouns than verbs. Here this finding was replicated but only for production – not for comprehension.

More data on the actual proportion of nouns and verbs in the input language are needed. Stoll et al. (2012) examine this but few other articles attempt this comparison. But given the similar proportions found on checklists in many different, unrelated languages, and the preponderance of nouns in early production, it seems likely that the composition of many checklists genuinely corresponds to the composition of early spoken vocabulary. This does not appear to have been a strategy in checklist composition but rather a product of the exhaustive methods generally used to construct the checklists (Dale and Penfold, 2011). Indeed, it might be problematic if those constructing checklists decided *a priori* that they must contain differing proportions of words in different word classes. Researchers should still not forget that the composition of early comprehension vocabulary is not the same as the composition of early production vocabulary.

## Contrasting Findings from Other Languages

### Production

There are a few studies that do not concur with these results. These include studies on Ngas, spoken in Nigeria, and on Mandarin.

Childers et al. (2007) suggested that the cultural context of child-rearing in Nigeria does not emphasize elicited

labeling or object-directed behavior. Here children's first words contained equal numbers of nouns and verbs. In rural Kenya, where caregivers are similarly often engaged in other activities and rarely participate in direct ostensive behavior with objects, older children are observed to attempt elicitation of all classes of words, and infants nevertheless still produced mainly nouns among their first spoken words.

Childers et al. (2007) suggest that children's verb learning may also be enhanced in Ngas due to features such as single syllable words and regular, rich verb inflection (carried on separate function words). Italian, Spanish and these Eastern Bantu languages have this rich verb inflection (Caselli et al., 1995; Bornstein et al., 2004) but still nouns predominate in early spoken words.

The combination of cultural and grammatical features in Ngas may together drive early production of verbs; though it is difficult to see why the same factors do not produce the same results in the Kenyan languages. One point to note is that the Childers et al. (2007) CDI had a smaller number of words than in most other inventories, and has no sound effects. Sound effects are a major category of children's early words, frequently used by both children and adults as spoken labels for objects (possibly due to auditory salience; Laing et al., 2016); in US English, Italian, and the Kenyan languages, children's first spoken words contain 20–30% sound effects.

Childers et al. (2007) also suggest that relevant verb features may be operating in Mandarin (Tardif et al., 1999). The Mandarin data though suffer from a scaling problem – the children learning Mandarin had relatively large spoken vocabularies, double that of the children in the same study learning English, and though the study scaled children's vocabulary, this leaves the composition of their vocabulary in doubt. Data from English and Dutch (Bornstein et al., 2004) do not demonstrate that monosyllabic verbs necessarily lead to early verb learning.

## Comprehension

Data from other languages concur with these findings that more verbs are comprehended early than are spoken. However, some researchers have doubted parents' abilities to report children's comprehension vocabularies accurately (Houston-Price et al., 2007), but other data suggest parents can report comprehension (Mills et al., 1993, 1997; Styles and Plunkett, 2009), including our data on individual words reported on this CDI (Alcock et al., 2015). The main issue with accuracy seems to be that parents find reporting overall vocabulary size easier than reporting the precise words children know, especially as vocabulary increases. Given consistency between studies and between languages, where methodology is constant, it is likely that parents are also relatively accurate in reporting the classes of word that children comprehend.

One argument for using parental report for comprehension at lower levels of vocabulary only is that parents may become confused once children's production vocabularies are larger. As children are less likely to produce verbs than nouns at lower levels of comprehension, parents may be more accurate in reporting the verbs. The structure of CDIs may also aid parents' recall of comprehension in low-production categories such as verbs, since words of one class are generally all clustered together on CDIs.

Pragmatic processes also explain why children comprehend more verbs than they produce. Goldfield (2000) suggests that caregiver structuring of interactions gives children opportunities to demonstrate and practice production of nouns but comprehension of verbs. Children in other sub-Saharan African cultures hear a reasonable proportion of commands (i.e., verb comprehension opportunities) in IDS, but also hear a wide range of other types of utterances (Nwokah, 1987; Kvalsvig et al., 1991; Rabain-Jamin, 1998; Deen, 2003). If Goldfield's explanation is valid, it implies that vocabulary knowledge may not differ between comprehension and production.

## Vocabulary Size

It is also helpful to consider whether children in this setting have comparable vocabulary levels to other settings, since verb/noun ratios depend on vocabulary size. In both production and comprehension mean vocabulary levels are intermediate between those found in UK English and those found in US English (Fenson et al., 1994; Hamilton et al., 2000). This is despite the extreme levels of poverty found in rural Kenya and the widely documented influence of poverty on early language and excess of children with language delay in low-income groups (see, for example Campbell et al., 2003).

## SUMMARY AND CONCLUSION

These data show that children hearing these two East African Bantu languages start by producing far more nouns than verbs but increase the proportion of verbs as their vocabulary increases. In contrast there is a more even distribution – and

no real change with age – between these two important word classes in comprehension. Kenyan children show some signs of learning verbs earlier than children learning to speak other languages, but there is no indication that verbs predominate in these children's first words as has been suggested for other languages (Brown, 1998; Tardif et al., 1999; Childers et al., 2007).

These findings imply that there may be no higher proportion of noun knowledge in early vocabulary, but simply a higher proportion of noun production. Explanations from pragmatics lend weight to this possibility. This has important implications for models of early word learning, including the ideas that nouns and/or object names are easier for children to learn. The factors that are hypothesized to assist in noun learning may still make nouns easier for children to produce, however.

The design of this study means that the data are comparable to those of Caselli et al. (1995) and to some extent to those of Bornstein et al. (2004). It is not possible to be as confident that the first words recorded here are genuinely comparable to those recorded by parents in the Tardif et al. (1999) study, where children's vocabularies were much larger. Likewise the composition of the vocabulary checklist in the Childers et al. (2007) study is not directly comparable to this or other previous studies.

An interesting related point is the relationship between age, vocabulary size, and vocabulary composition. The Mandarin- and English-learning children in the Tardif et al. (1999) study were of the same age but different vocabulary sizes. In Bornstein et al.'s (2004) cross-linguistic study vocabulary was recorded for all of the children at the same age, while in this study and Caselli et al.'s (1995) study children were younger and of a variety of ages, but some of the children had comparable vocabulary sizes to those in Bornstein's study. However, there are some indications that children with the same vocabulary size, speaking the same language, but of different ages, may have different vocabulary compositions (Rowland et al., 2016).

While studying this phenomenon in these languages is interesting in that little is known about vocabulary development in Eastern Bantu languages nor in children growing up in sub-Saharan African cultures, our study is not just of interest for this reason. Using an internationally accepted method of studying early language comprehension and production, but in understudied languages and a non-WEIRD (Henrich et al., 2010) setting, makes our findings – confirming and extending previous studies – additionally valid and, it can be argued, more interesting.

Many previous studies examining noun and verb learning in early language have not collected data on comprehension. The comparison here with English and Italian represents one of the few published studies of directly comparable data, with enough detail within the published article, to enable a direct comparison. A future larger-scale study such as that of Bornstein et al. (2004), but concentrating on younger children and either collecting additional data on comprehension, or utilizing one of the publicly

available CDI datasets (Frank et al., 2017), could therefore be highly informative. The composition of vocabulary scales must though be directly comparable (avoiding issues such as the elimination of large, important early categories of vocabulary as in Childers et al., 2007), and the composition of the actual input language to children's should also be a priority (Stoll et al., 2012).

## ETHICS STATEMENT

The Kenya Medical Research Institute National Scientific and Ethical Committees approved the study (SCC No: 832). Informed consent was obtained from all families and guardians of study participants. Because of the nature of the sample, and the number of illiterate parents, consent was obtained orally from many participants.

## REFERENCES

- Alcock, K. J., Rimba, K., Abubakar, A., Holding, P. A., Kitsao-Wekulo, P. K., and Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: CDI construction in for two Kenyan languages. *J. Child Lang.* 42, 763–785. doi: 10.1017/S0305000914000403
- Alcock, K. J., Rimba, K., and Newton, C. R. (2012). Early production of the passive in two Eastern Bantu languages. *First Lang.* 32, 459–478. doi: 10.1177/0142723711419328
- Altinkamış, N. F., Kern, S., and Sofu, H. (2014). When context matters more than language: verb or noun in French and Turkish caregiver speech. *First Lang.* 34, 537–550. doi: 10.1177/0142723714560179
- Au, T. K.-F., Dapretto, M., and Song, Y. K. (1994). Input vs. constraints: early word acquisition in Korean and English. *J. Mem. Lang.* 33, 567–582. doi: 10.1006/jmla.1994.1027
- Bassano, D. (2000). Early development of nouns and verbs in French: exploring the interface between lexicon and grammar. *J. Child Lang.* 27, 521–559. doi: 10.1017/S0305000900004396
- Bates, E. (1979). *The Emergence of Symbols - Cognition and Communication in Infancy*. New York, NY: Academic Press.
- Bornstein, M. H., and Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities. *Infancy* 7, 299–316. doi: 10.1207/s15327078in0703\_5
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Dev.* 75, 1115–1139. doi: 10.1111/j.1467-8624.2004.00729.x
- Brown, P. (1998). Children's first verbs in Tzeltal: Evidence for an early verb category. *Linguistics* 36, 713–754. doi: 10.1515/ling.1998.36.4.713
- Camaioni, L., Longobardi, E., Venuti, P., and Bornstein, M. H. (1998). Maternal speech to 1-year-old children in two Italian cultural contexts. *Early Dev. Parent.* 7, 9–18. doi: 10.1002/(SICI)1099-0917(199803)7:1<9::AID-EDP159>3.0.CO;2-T
- Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., et al. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Dev.* 74, 346–357. doi: 10.1111/1467-8624.7402002
- Caselli, M., Bates, E., Casadio, P., and Fenson, J. (1995). A cross-linguistic study of early lexical development. *Cogn. Dev.* 10, 159–199. doi: 10.1016/0885-2014(95)90008-X
- Childers, J. B., Vaughan, J., and Burquest, D. A. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. *J. Child Lang.* 34, 199–225. doi: 10.1017/S0305000906007835
- Choi, S., and Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *J. Child Lang.* 22, 497–529. doi: 10.1017/S0305000900009934

## AUTHOR CONTRIBUTIONS

KA designed and implemented the study, supervised data collection, carried out the analysis and wrote the manuscript.

## FUNDING

Data collection was supported in part by NIMH Fogarty R21 award (Grant MH72597-02).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01898/full#supplementary-material>

- Dale, P. S., and Penfold, M. (2011). *Adaptations of the MacArthur-Bates CDI into non-U.S. English Languages*. Available at: <http://www.sci.sdsu.edu/cdi/documents/AdaptationsSurvey7-5-11Web.pdf>
- Deen, K. U. (2003). *The Acquisition of Nairobi Swahili: The Morphosyntax of Inflectional Prefixes and Subjects*. Ph.D. thesis, University of California, Los Angeles, CA.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., and Pethick, S. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1–173; discussion 174–85. doi: 10.2307/1166093
- Fernald, A., and Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Dev.* 64, 637–656. doi: 10.1111/j.1467-8624.1993.tb02933.x
- Frank, M. C., Braginsky, M., Yurovsky, D., and Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *J. Child Lang.* 44, 677–694. doi: 10.1017/S0305000916000209
- Gentner, D. (1982). "Why nouns are learned before verbs: Linguistic relativity versus natural partitioning," in *Language Development: Language, Thought, and Culture*, Vol. 2, ed. S. A. Kuczaj (Hillsdale, NJ: Erlbaum), 301–334.
- Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Lang. Learn. Dev.* 1, 23–64. doi: 10.1207/s15473341l1d0101\_4
- Goldfield, B. A. (2000). Nouns before verbs in comprehension vs. production: the view from pragmatics. *J. Child Lang.* 27, 501–520. doi: 10.1017/S0305000900004244
- Goldfield, B. A., and Reznick, J. S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *J. Child Lang.* 17, 171–183. doi: 10.1017/S0305000900013167
- Goldin-Meadow, S., Seligman, M. E. P., and Gelman, R. (1976). Language in the two-year old. *Cognition* 4, 189–202. doi: 10.1016/0010-0277(76)90004-4
- Hamadani, J. D., Baker-Henningham, H., Tofail, F., Mehrin, F., Huda, S. N., and Grantham-McGregor, S. M. (2010). Validity and reliability of mothers' reports of language development in 1-year-old children in a large-scale survey in Bangladesh. *Food Nutr. Bull.* 31, S198–S206. doi: 10.1177/15648265100312S212
- Hamilton, A., Plunkett, K., and Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *J. Child Lang.* 27, 689–705. doi: 10.1017/S0305000900004414
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). Most people are not WEIRD. *Nature* 466, 29–29. doi: 10.1038/466029a
- Houston-Price, C., Mather, E., and Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behaviour in a preferential looking task. *J. Child Lang.* 34, 701–724. doi: 10.1017/S0305000907008124
- Kauschke, C., and Hofmeister, C. (2002). Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life. *J. Child Lang.* 29, 735–757. doi: 10.1017/S0305000902005330

- Kim, M., McGregor, K. K., and Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: language-general and language-specific patterns. *J. Child Lang.* 27, 225–254. doi: 10.1017/S0305000900004104
- Kvalsvig, J., Liddell, C., Reddy, A., Qotyana, P., and Shabalala, A. (1991). Communication and teaching in the home: a study of Zulu and Sotho preschoolers. *Early Child Dev. Care* 74, 61–81. doi: 10.1080/0300443910740105
- Laing, C. E., Vihman, M., and Keren-Portnoy, T. (2016). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *J. Child Lang.* 44, 1117–1139. doi: 10.1017/S0305000916000428
- Markman, E. M. (1990). Constraints children place on word meanings. *Cogn. Sci.* 14, 57–77. doi: 10.1207/s15516709cog1401\_4
- McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., and Lannon, R. (2011). An image is worth a thousand words: why nouns tend to dominate verbs in early word learning. *Dev. Sci.* 14, 181–189. doi: 10.1111/j.1467-7687.2010.00968.x
- Mills, D. L., Coffey-Corina, S., and Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 to 20 months. *Dev. Neuropsychol.* 13, 397–445. doi: 10.1080/87565649709540685
- Mills, D. L., Coffey-Corina, S. A., and Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *J. Cogn. Neurosci.* 5, 317–334. doi: 10.1162/jocn.1993.5.3.317
- Nwokah, E. E. (1987). Maidese versus motherese-is the language input of child and adult caregivers similar? *Lang. Speech* 30, 213–237. doi: 10.1177/002383098703000303
- Rabain-Jamin, J. (1998). Polyadic language socialization strategy: the case of toddlers in Senegal. *Discourse Process.* 26, 43–65. doi: 10.1080/01638539809545037
- Rowland, C. F., Meints, K., Just, J., Brelsford, V., Christopher, A., and Alcock, K. J. (2016). The effect of age on the composition of children's first 10 words: Evidence from the UK-CDI. *Paper Presented at the International Conference on Infant Studies*, New Orleans.
- Schafer, G. (2005). Infants can learn decontextualized words before their first birthday. *Child Dev.* 76, 87–96. doi: 10.1111/j.1467-8624.2005.00831.x
- Simons, G. F., and Fennig, C. D. (2017). *Ethnologue: Languages of the World*. Available at: <http://www.ethnologue.com/>
- Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., et al. (2012). Nouns and verbs in Chintang: children's usage and surrounding adult speech. *J. Child Lang.* 39, 284–321. doi: 10.1017/S0305000911000080
- Styles, S., and Plunkett, K. (2009). What is 'word understanding' for the parent of a one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *J. Child Lang.* 36, 895–908. doi: 10.1017/S0305000908009264
- Tamis-LeMonda, C. S., Kuchirko, Y., and Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Curr. Dir. Psychol. Sci.* 23, 121–126. doi: 10.1177/0963721414522813
- Tardif, T., Fletcher, P., Liang, W. L., Zhang, Z. X., Kaciroti, N., and Marchman, V. A. (2008). Baby's first 10 words. *Dev. Psychol.* 44, 929–938. doi: 10.1037/0012-1649.44.4.929
- Tardif, T., Gelman, S. A., and Xu, F. (1999). Putting the "noun bias" in context: a comparison of English and Mandarin. *Child Dev.* 70, 620–635. doi: 10.1111/1467-8624.00045
- Tardif, T., Shatz, M., and Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin. *J. Child Lang.* 24, 535–565. doi: 10.1017/S030500099700319X
- Xuan, L., and Dollaghan, C. (2013). Language-specific noun bias: evidence from bilingual children. *J. Child Lang.* 40, 1057–1075. doi: 10.1017/S0305000912000529

**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 © 2017 Alcock. 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.



# When Meaning Is Not Enough: Distributional and Semantic Cues to Word Categorization in Child Directed Speech

Sara Feijoo<sup>1\*</sup>, Carmen Muñoz<sup>1</sup>, Anna Amadó<sup>2</sup> and Elisabet Serrat<sup>2</sup>

<sup>1</sup> Department of Modern Languages and English Studies, University of Barcelona, Barcelona, Spain, <sup>2</sup> Department of Psychology, University of Girona, Girona, Spain

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Chloe Marshall,  
UCL Institute of Education,  
United Kingdom  
Mirta Vernice,  
University of Milano-Bicocca, Italy

### \*Correspondence:

Sara Feijoo  
sfeijoo@ub.edu

### Specialty section:

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

**Received:** 10 March 2017

**Accepted:** 06 July 2017

**Published:** 19 July 2017

### Citation:

Feijoo S, Muñoz C, Amadó A and  
Serrat E (2017) When Meaning Is Not  
Enough: Distributional and Semantic  
Cues to Word Categorization in Child  
Directed Speech.  
Front. Psychol. 8:1242.  
doi: 10.3389/fpsyg.2017.01242

One of the most important tasks in first language development is assigning words to their grammatical category. The *Semantic Bootstrapping Hypothesis* postulates that, in order to accomplish this task, children are guided by a neat correspondence between semantic and grammatical categories, since nouns typically refer to objects and verbs to actions. It is this correspondence that guides children's initial word categorization. Other approaches, on the other hand, suggest that children might make use of distributional cues and word contexts to accomplish the word categorization task. According to such approaches, the *Semantic Bootstrapping* assumption offers an important limitation, as it might not be true that all the nouns that children hear refer to specific objects or people. In order to explore that, we carried out two studies based on analyses of children's linguistic input. We analyzed child-directed speech addressed to four children under the age of 2;6, taken from the CHILDES database. The corpora were selected from the Manchester corpus. The corpora from the four selected children contained a total of 10,681 word types and 364,196 word tokens. In our first study, discriminant analyses were performed using semantic cues alone. The results show that many of the nouns found in parents' speech do not relate to specific objects and that semantic information alone might not be sufficient for successful word categorization. Given that there must be an additional source of information which, alongside with semantics, might assist young learners in word categorization, our second study explores the availability of both distributional and semantic cues in child-directed speech. Our results confirm that this combination might yield better results for word categorization. These results are in line with theories that suggest the need for an integration of multiple cues from different sources in language development.

**Keywords:** semantic cues, distributional cues, word categorization, child-directed speech, grammatical categories

## INTRODUCTION

One of the most significant challenges for children when learning their first language is assigning words to their corresponding syntactic categories. For instance, how do English-learning children know that 'table' is a noun, 'eat' is a verb, and 'kiss' can be both a noun and a verb? Generativist approaches have put forward the so-called *Semantic Bootstrapping Hypothesis* (Pinker, 1984;

Fodor, 1998; Laurence and Margolis, 2001), which predicts that children use semantic information to map words into their corresponding grammatical category. In particular, children are said to have innately specified information in terms of nouns referring to objects, and verbs referring to actions. The present paper undertakes a critical examination of these assumptions: on the one hand, it tests the reliability of semantic information by examining the amount of nouns in children's input that refer to specific objects or people; on the other hand, it examines the accuracy with which words could be categorized on the basis of a combination of multiple cues (i.e., both semantic and distributional cues).

The strength of the semantic bootstrapping approach lies on the idea that mappings between semantics and grammatical classes are universal (e.g., nouns denote objects while verbs denote actions in any language). On the contrary, other types of cues like phonological or distributional cues are language-specific. In particular, the noun-object mapping seems easily attainable in language development, as many studies have highlighted a noun bias in children's early vocabularies across different languages, mainly because of the conceptual simplicity that nouns exhibit (Jackson-Maldonado et al., 1993; Bates et al., 1994; Caselli et al., 1999; Bassano, 2000; Gleitman et al., 2005). Concreteness or imageability might be the underlying predictor of the identifiability of nouns from their observed extralinguistic contexts. Thus, learners' noun bias might be based on the assumption that object-reference items are the best ones that fit in a word-to-world pairing procedure. With the meaning of nouns, and the intuition that nouns relate to real-world objects, children might then start building a rudimentary nominal grammatical category.

Then, the semantic bootstrapping proposal claims that one source of information about the meaning of words is available from the beginning of the language learning process, and it constitutes the basis from which learners start building their initial grammatical categories. This initial information source allows the learner to acquire a subset of lexical items (i.e., nouns that refer to specific objects) which requires little linguistic knowledge and is pragmatically supported. Thus, the grounding of grammatical categories would start from the identification of semantic categories first, and semantic features would later on bootstrap grammar.

Nevertheless, the main problem within the semantic bootstrapping approach is that it presupposes that correlations between semantics and grammatical categories are perfect mappings in any language. Furthermore, such an approach also assumes that children start the language learning process with the expectation that such mappings actually exist. However, this assumed straightforward concept-word pairing is somehow problematic, as some have pointed out (Ambridge and Lieven, 2011).

To start with, naive learners with no prior knowledge of grammatical categories in their language and who are exposed to fluent speech might even fail to perform the object referent-noun mapping. Even if child learners were ready to map external referents to particular words in the input, how does the child

know which word does the "object" semantic component refer to, out of the possible words she hears? First of all, the words to which children are exposed might refer to objects which are absent from the child's sight when they are spoken (Gleitman, 1990). In addition, when addressing children, parents may refer to the same object with different words in different contexts (Yurovsky et al., 2012). Furthermore, the mapping task becomes even less clear when facing multiple-word utterances (Yu and Ballard, 2007). However, such utterances are the ones which children are most likely to encounter in the course of their linguistic development, as only a small percentage of utterances in child-directed speech have been reported to contain words in isolation (Bernstein Ratner and Rooney, 2001; Monaghan and Christiansen, 2010; Monaghan and Mattock, 2012; Feijoo and Hilferty, 2013). Furthermore, research shows that mothers never or hardly ever use words in isolation, even in situations where they are explicitly teaching new vocabulary to their children (Aslin et al., 1996).

A further problem concerns the way the object-word mapping itself should proceed. It is known as *Quine's Gavagai problem*, or the problem of referential indeterminacy (Quine, 1960). Imagine a mother-child interaction situation where the mother would point to a brown running dog and say "Look at the dog!". Even if the visual image and the target word *dog* were immediately associated, how does the child know that the word *dog* actually refers to the dog itself and not, say, to a more general referent such as *animal*, or to a specific type of dog, or to a part of the dog (i.e., its legs, its tail...), or to a physical property of the dog (i.e., its colour, its fur...), or to the action of running itself? How do children know that they can equally use the word *dog* to refer to another type of dog (i.e., a sitting dog, a white dog, a dog from a different breed, etc.) and they cannot use it with other animals like a brown running cat? A mere world-to-word mapping assumption cannot account for children's choice and learning of the word *dog* and its natural referents in the real world.

Recent empirical evidence has shown that cross-situational statistical learning might be the key to solve the referential ambiguity problem illustrated in the *Gavagai problem* (Smith and Yu, 2008; Scott and Fisher, 2012; Vlach and Johnson, 2013; Suanda et al., 2014; Benitez et al., 2016). This statistical learning approach suggests that adults as well as young children are able to map linguistic units to referents in the world by tracking co-occurrence probabilities across different learning situations. Thus, it seems that learners are sensitive to the statistical consistency with which a given word is used in front of a given referent and the mapping between heard word and seen referent can occur in this way.

However, the noun-object mapping proposal also assumes that the set of nouns which children hear from their input specifically refer to objects in the real world. What if children were exposed to superordinate terms? Or what if they were exposed to abstract words whose meaning does not relate to a specific object or referent? Traditional linguistic analyses postulate that the category "noun" is both a notional and a grammatical concept (Lyons, 1977). There is a central semantic concept of noun, which

is present in all languages, and which includes words for persons, animals, and things. All the other more abstract ontological categories denoted by nouns appear to be generalizations from this core concept.

Acknowledging that this core concept includes only a subset of all possible nouns gives rise to the question of how large the proportion of nouns belonging to this subset actually is in English. In other words, can this subset account for all the examples of nouns that English-learning children are exposed to? Previous studies have already pointed out that semantic criteria alone do not provide a reliable basis to determine the category membership of many words in English, since there are many nouns which do not denote physical objects (e.g., *an explanation*) (Yu and Ballard, 2007; Tare et al., 2008), or there are many words in English which can be both classified as nouns or verbs (e.g., *a kiss* vs. *to kiss*, *a walk* vs. *to walk*, etc.) (Nelson, 1995; Maratsos, 1999; Tomasello, 2010; Conwell and Morgan, 2012).

In this line, Nelson et al. (1993) propose two distinguishable semantic classes of nouns: on the one hand, BLOCS neatly correspond to basic level object categories; on the other hand, XBLOCS refer to all those words which would naturally fall out of the cognitive basic level, either because their extralinguistic referent is too general or because they do not refer to a specific object referent at all. Nelson et al. (1993: p. 71) further distinguished different types of XBLOCS taking their meaning into account:

- *Locations*: places, both indoors and out (e.g., *beach*, *kitchen*)
- *Actions*: single actions, specific or general (e.g., *kiss*, *help*)
- *Superordinate/generic*: terms that denote what are generally considered superordinate categories (e.g., *toys*, *animals*).
- *Events*: terms that refer to complex events that take place through time (e.g., *lunch*, *party*).
- *Person roles*: roles that people play in social/cultural life (e.g., *doctor*, *brother*).
- *Natural phenomena*: states, actions, events or entities (e.g., *sky*, *snow*, *clouds*).
- *Temporal entities*: e.g., *morning*, *day*.
- *Quantities*: e.g., *drop*.

In their study, Nelson et al. (1993) showed that children learn and use many XBLOC words early in the language learning process. As multi-word combinations and productivity in noun morphology develops, noun roles are assigned to these words, showing that they are accurately categorized as nouns. Therefore, words which lack the expected semantic content that would yield their successful categorization (and which would therefore be left unclassified on the basis of semantics alone) are nonetheless accurately classified in their right word class from the early stages of the nominal category building process.

Other studies also point out that object nouns or action verbs do not necessarily dominate children's earliest lexical productions, neither in English, nor in other languages (Gopnik and Choi, 1995; Bassano, 2000). And still, those words which do not neatly map into their corresponding semantic category are correctly classified by children and they are used in a grammatical way.

Given the evidence provided, it seems clear that there are other types of cues at work, alongside semantics, when words are being classified into their corresponding grammatical category. For instance, at least as far as English is concerned, several studies have provided evidence for the usefulness of phonological information as a key element for the access to grammatical properties of the language (e.g., Kelly, 1996; Monaghan et al., 2005, 2007; Fitneva et al., 2009). These findings suggest that, on the one hand, these phonological cues are reliably found in parents' language. On the other hand, evidence has also been found that young language learners as well as adults are aware of such cues and their correlation to grammatical categories.

Besides, syntactic or distributional information might be a very powerful cue that assists young language learners in language development as well. Regarding word categorization, the context of a word with respect to other words in the same sentence might provide indications about the category of that word in English. For example, English nouns are typically preceded by determiners and followed by nominal morphology (e.g., *the babies*), while verbs are typically preceded by auxiliaries or strong subject pronouns and followed by verbal morphology (e.g., *she walked*).

Studies on computer simulations have provided evidence for the usefulness of distributional and positional information for an initial categorization of words in the absence of semantic or referential information (Cartwright and Brent, 1997; Redington et al., 1998). Such distributional information appears to be available not only to adult speakers but also to young language learners as well (Mintz, 2003; Monaghan et al., 2007; Feijoo et al., 2015).

Furthermore, empirical evidence from artificial language studies seems to suggest that children's learning of grammatical structure as well as their word-reference associations improve when words are coherently marked by a combination of different types of cues, either phonological, distributional or semantic cues (Gomez and Lakusta, 2004; Gerken et al., 2005; Lany and Gomez, 2008; Lany, 2014). However, when words are not reliably marked by these cues, infants fail to learn their semantic or grammatical properties, since young language learners are more likely to learn from deterministic rather than probabilistic cues, and they would only rely on relatively robust correlations between word-forms and their corresponding grammatical category (Lany and Saffran, 2010; Yurovsky et al., 2012; Lany, 2014).

In particular, Lany and Saffran (2010) found that experience with reliable distributional cues in the input is a key factor that predicts children's learning of word meanings: when words' distributional properties correctly indicated the grammatical category to which words belonged, infants successfully learned word-referent mappings. In contrast, infants failed at the word-referent pairing task when distributional cues were not reliably correlated with grammatical category membership.

The main goal of the present study was to test the likelihood with which children could classify all the nouns they hear in their corresponding grammatical category using semantic cues derived from the input. To this end, two studies were carried out: in the first one, a corpus-based analysis of child-directed speech explores the potential strength of semantic information alone in

**TABLE 1** | Characteristics of the selected corpora.

|              | Anne              | Aran              | Becky            | Carl             |
|--------------|-------------------|-------------------|------------------|------------------|
| Age range    | 1;10.07 to 2;6.29 | 1;11.12 to 2;6.17 | 2;0.07 to 2;6.29 | 1;8.22 to 2;6.19 |
| Sex          | Female            | Male              | Female           | Male             |
| Total types  | 2,761             | 3,432             | 2,222            | 2,266            |
| Total tokens | 103,457           | 118,469           | 65,411           | 76,859           |

children's input. The second study examines the benefits that a combination of semantic and distributional information could provide to language learning children when facing the task of word categorization.

## STUDY 1

### Objective

While it is not still clear whether the first analysis that children perform on the input is on notional grounds (and therefore, semantic cues are considered first) or distributional grounds (and thus, syntactic cues are considered first), it is widely accepted that the semantic notion of *object* and *action* might assist language learners in the identification of nouns and verbs, respectively. As mentioned earlier, one of the problems for the *Semantic Bootstrapping* proposal (Pinker, 1984) lies on the difficulty of identifying the meaning of unknown words and, consequently, their semantic category. Besides, the links between semantic categories and grammatical categories are not one-to-one but many-to-many (Mintz, 2003). Thus, for example, not all items within the semantic category of actions are verbs. An adjective such as *noisy*, or a noun such as *call*, can also be semantically classified as actions. In fact, as Nelson et al. (1993) have pointed out, the actual proportion of English nouns which conform to the semantic category of objects is only a small subset of the whole noun inventory.

Acknowledging that the core traditional definition of nouns as labels for people, animals, and things only includes a subset of all English nouns raises the question of what is the actual percentage of nouns which belong to this subset (i.e., how big the subset is, considering all English nouns). It also raises the issue of whether this smaller subset can be taken to account for all of the nouns that young children are exposed to and will later acquire.

Thus, the main objective of this first study is to examine the usefulness and reliability of semantic cues for the categorization of nouns in English child-directed speech. While it is true that not all English nouns refer to objects exclusively, it is also true that the kind of interactions that very young children are involved in are often restricted to the *here* and *now* and to familiar objects within each child's reach (Baldwin, 1993; Clark, 2009). A close evaluation to words which semantically refer to objects will be performed in order to test how large the set of object-referring nouns actually is in children's input. Furthermore, words which semantically refer to actions will also be analyzed in order to examine the possible overlap between the grammatical categories of nouns and verbs which have semantic content in common.

**TABLE 2** | Characteristics of the groups of lexical items from all the corpora.

|                              | Total types | Total tokens | TTR   |
|------------------------------|-------------|--------------|-------|
| Total all corpora            | 10,681      | 364,196      | 0.029 |
| Total selected lexical items | 9,621       | 139,624      | 0.069 |
| Total nouns                  | 5,397       | 51,577       | 0.104 |
| Total non-nouns              | 4,233       | 88,047       | 0.048 |

### Corpus Preparation

We analyzed child-directed speech addressed to four children under the age of 2;6, taken from the CHILDES database (MacWhinney, 2000). The corpora were selected from the Manchester corpus (Theakston et al., 2001) and included the files from Aran, Carl, Anne, and Becky. The characteristics of the selected corpora are summarized in **Table 1**.

All the lexical items from each corpus were then classified into two different categories to be analyzed separately. One category included all nouns, and the other category, the "non-noun" category, included all verbs, adjectives, and adverbs. For dual-class words, that is, English words that can, for instance, be both classified as nouns and verbs (e.g., *kiss*, *call*, *brush*), the KWAL utility of the CLAN program was used in order to work out the exact number of tokens that were used as nouns and the number of tokens that were used as verbs in every transcript. **Table 2** shows a summary of the characteristics of each group, with the total number of types and the total number of tokens found in each group. It also shows the Type/Token ratio as an indicator of lexical diversity. However, note that such indicator will not be a variable considered in our analyses.

### Cue Derivation

All the lexical items from the child corpora were further classified into different groups, according to the semantic features they bore. The main goal was to analyze the consistency and reliability with which the relationship between semantic information and grammatical categories is represented in the input addressed to English-learning children. Particular attention was paid to the semantic overlap between nominal elements which describe actions and prototypical verbal elements which equally describe actions. Such contradictory information might be especially misleading for any child who relies on semantic information to form grammatical categories, since they would wrongly classify action nouns as verbs.

A set of semantic cues which have been said to identify nouns (Nelson et al., 1993) was selected. Not only the group of nouns, but also the set of verbs, adjectives, and adverbs

from the “non-noun” group in the selected corpora, were tested against the selected cues. This made it possible to analyze the degree of overlap between semantic and grammatical categories and, therefore, to work out the risk of misclassifying elements into their wrong grammatical category on the basis of semantic information. Following the work by Nelson et al. (1993), the set of semantic cues that were selected for the analysis include the following:

- *Sem1*: this was the group for proper nouns, which semantically refer to single individuals and never expand to whole-class reference. A number of classical studies (e.g., Katz et al., 1974; Gelman and Taylor, 1984) have already provided evidence for the fact that children understand proper nouns and common nouns differently from very early stages of language development. Proper nouns of people, animals, toys, stories, songs, places, holidays, etc., scored 1 in this category and 0 everywhere else.
- *Sem2*: this was the category that best corresponded with the traditional notional definition of nouns as labels for names of people, animals, and objects. In particular, the *Sem2* group was meant to include all common count nouns that could be considered members of the basic level category in Rosch’s (1978) terms (e.g., *dog, apple, chair*). Words from the corpora which matched this definition scored 1 here and they scored 0 everywhere else.
- *Sem3*: this was the category that included all mass nouns in the corpora (e.g., *milk, paper*). They differ from the nouns from the *Sem2* category in the sense that *Sem3* words do not denote discrete entities but whole substances. This difference is also reflected in syntax, since mass nouns do not combine with the same type of determiner as count nouns. Furthermore, children seem to be aware of such differences by their second year of life (Soja, 1992). Words that matched the *Sem3* description scored 1 here and they scored 0 everywhere else.
- *Sem4*: this was the category which Nelson et al. (1993) labeled as XBLOCs (i.e., not basic level object categories). These were words which could not be included in the *Sem3* category or in the *Sem2* category. While Nelson et al. (1993) distinguish several different types of XBLOC nouns, for the purpose of the present study only two different groups of the *Sem4* category were made:
  - *Sem4a*: this was the group for words that described actions (e.g., *kiss, help, trip*). Nouns that matched this description scored 1 here and 0 in the other *Sem* categories. Crucially, verbs which matched this description also scored 1 here.
  - *Sem4b*: this group included all the other XBLOC nouns from the study by Nelson et al. (1993). Among them were words which denoted locations and places, both indoors and outdoors (e.g., *kitchen, park, school*); words which described generic categories and belonged to the superordinate level in Rosch’s (1978) terms (e.g., *animal, thing*); words which referred to abstract events or social gatherings (e.g., *lunch, party*); words which described person roles (e.g., *doctor, brother*); words which denoted natural phenomena (e.g.,

*sky, heat*); words which referred to temporal entities (e.g., *morning, day*); and words which designated quantities (e.g., *drop, spoonful*). Then, words that semantically referred to any of these groups scored 1 in this category and 0 in all the other *Sem* categories.

What motivates the division of the *Sem4* group into two subgroups is that, unlike Nelson et al. (1993), the purpose of our analysis is not to provide an accurate description and classification of XBLOC nouns in general. Instead, the main objective is to test the amount of English nouns that lack a direct semantic component available to language learners (i.e., which nouns are XBLOC nouns and which ones are not). Besides, it was also important to analyze the degree of semantic overlapping between nouns and other words such as verbs, for which the *Sem4a* category was created.

In order to guarantee reliability on coding, words were classified into their corresponding semantic categories by two different raters. Inter-rater reliability was measured using the Cohen’s Kappa statistic ( $Kappa = 0.926, p < 0.05$ ).

## Results and Discussion

The set of nouns and the set of non-nouns which were previously obtained were tested using the four different semantic cues described above. The total number of nouns that met each of the semantic descriptions under consideration is shown in **Table 3**. In a similar way, **Table 4** shows the results obtained from the equivalent analysis with the rest of open class words. As shown in the table, while most non-nouns (i.e., a total of 2,520) belong to the *Sem4a* group, since they denote actions, the remaining 1,654 non-nominal types were not captured by any of the semantic features under consideration.

Correct classification of all types and tokens with all the semantic cues was tested with discriminant analyses. A discriminant analysis is a multivariate inferential technique. Its main objective is to classify individuals in two groups according to a number of previously selected variables. It works out the reliability with which the variables accurately describe the

**TABLE 3 |** Total number of nouns in each semantic category.

|                                         | Total types | Total tokens | TTR   |
|-----------------------------------------|-------------|--------------|-------|
| Proper nouns ( <i>Sem1</i> )            | 699         | 9,516        | 0.073 |
| Basic level count nouns ( <i>Sem2</i> ) | 2,675       | 25,369       | 0.105 |
| Basic level mass nouns ( <i>Sem3</i> )  | 288         | 2,655        | 0.108 |
| Action words ( <i>Sem4a</i> )           | 228         | 1,419        | 0.161 |
| Non-basic level words ( <i>Sem4b</i> )  | 1,507       | 12,618       | 0.119 |

**TABLE 4 |** Total number of non-nouns in each semantic category.

|                                         | Total types | Total tokens | TTR   |
|-----------------------------------------|-------------|--------------|-------|
| Proper nouns ( <i>Sem1</i> )            | 0           | 0            | –     |
| Basic level count nouns ( <i>Sem2</i> ) | 0           | 0            | –     |
| Basic level mass nouns ( <i>Sem3</i> )  | 0           | 0            | –     |
| Action words ( <i>Sem4a</i> )           | 2,520       | 46,530       | 0.054 |
| Non-basic level words ( <i>Sem4b</i> )  | 59          | 1,318        | 0.045 |

members of a given group and whether presence or absence of a given set of variables determines group membership. Regarding types, when the variables *Sem1*, *Sem2*, *Sem3*, and *Sem4a* were entered simultaneously, overall correct classification reached 82.0%, Wilks  $\lambda = 0.439$ ,  $\chi^2 = 7916.359$ ,  $p < 0.001$ . However, this high overall correct classification was obtained mainly because of the high score obtained in the correct classification of non-noun words, which was 98.7%. This indicates that there were almost no verbs, adjectives or adverbs which carried any of the semantic features which are typically associated to nouns. However, correct classification among nouns lowered to 67.9%.

The same analysis that was run with types was also run with all tokens from the four corpora. For the token analysis, when the same semantic cues were entered simultaneously, only 39.9% of nouns were correctly classified. Overall correct classification was 66.3% of tokens, Wilks  $\lambda = 0.945$ ,  $\chi^2 = 544.696$ ,  $p < 0.001$ . As with types, overall correct classification was relatively high because of the high scores in the non-noun group, with identical results in the token analysis as in the type analysis. However, such high scores in the non-noun group only indicate that the group of nouns which could be potentially created on the basis of semantic information is very accurate (i.e., there are very few non-noun words which are at risk of being misclassified as nouns on the basis of their semantic content). On the contrary, completeness scores (i.e., the number of nouns which are correctly classified as such) are relatively low, which indicates that only a subset of nouns would be correctly classified as nouns on the basis of their meaning, and most nouns would be wrongly classified as non-nouns.

These analyses show that there is an important number of noun tokens which lack the semantic features with which nouns are associated, and whose semantic information is either ambiguous or too broad. Thus, children would not be able to work out the grammatical category to which these nouns belong on the basis of semantic information alone. On the contrary, other sources of information might be necessary for the correct categorization of most of the nouns to which children are exposed. Our second study tested the likelihood with which semantic and syntactic cues together could yield better results than semantic cues alone for word categorization.

## STUDY 2

### Objective

As seen from the results of Study 1, the semantic notion of “object” only correlates with the grammatical category of nouns in a very weak way. The purpose of this second study is, then, to find a second source of information which, alongside with semantic information, might assist young English learners in the categorization of the nouns they hear from the input. In this line, Yu (2006) found that the association between words and objects was assisted by the presence of syntactic information. Furthermore, analyses of child-directed speech corpora highlight the usefulness of multiple cue integration for word categorization (Monaghan et al., 2007; Monaghan and Mattock, 2012; Yurovsky et al., 2012). In more general terms, several studies also highlight

the importance of redundant information in language learning (e.g., Gogate et al., 2000; Frank et al., 2009; Smith et al., 2010; Riordan and Jones, 2011). Thus, combined cues seem to provide better language learning outcomes.

As mentioned earlier in the introduction, recent findings from artificial language experiments (Lany and Saffran, 2010; Lany, 2014) also suggest that the presence of robust correlations between the distributional and semantic properties of words enhances infants’ word learning and word categorization. Thus, the objective of this second study was to test how robust these correlations between semantic and distributional cues actually are in natural child-directed speech, given that it significantly differs from artificial languages in terms of complexity and the presence of “noisy” elements.

### Corpus Preparation

For the second study, we used the same corpora as for Study 1. We also followed the same procedure and criteria regarding the classification of lexical items into the “noun” and the “non-noun” groups.

### Cue Derivation

The same semantic cues that were used in Study 1 were also used in Study 2. Regarding distributional cues, following previous studies (Mintz, 2003; Monaghan et al., 2007; Feijoo et al., 2015), for the present analysis we only considered the set of syntactic contexts which included extremely local grammatical relationships of the type *Determiner + Noun* (i.e., English articles, demonstrative determiners, possessive determiners, and quantifiers preceding nouns).

A list of six different distributional contexts was generated and every target word was analyzed to see whether its context matched any of the six established. Words scored 1 if they appeared in any of those syntactic contexts and they scored 0 otherwise. In this sense, this analysis with distributional cues was slightly different from the one using semantic cues. In terms of semantic cues, when a word (type or token) scored 1 in a given category, it scored 0 in all other categories as well (e.g., a word cannot be a proper noun (category *Sem1*) and a common object (category *Sem2*) at the same time). The same is true of the Token analysis using distributional cues. However, as far as the Type analysis using distributional cues is concerned, it is not true that when a Type scores 1 in a given distributional context it scores 0 in all the other contexts as well. For instance, we found occurrences of the phrase *a dog* and the phrase *the dog* in the corpus of the same child. In terms of tokens, both instantiations of the word *dog* correspond to two different tokens, with their respective different syntactic context each. However, in terms of types, the same word type *dog* is both found in two different syntactic contexts. That is why, as a Type, *dog* scored 1 in both contexts at the same time (for further examples see the Supplementary Material, with an excerpt of the classified material).

In order to obtain the different distributional contexts in which every word in our corpus occurred, the COOCCUR utility of the CLAN program was used to generate a list of every word in our corpus plus the word which occurred immediately before it,

as well as the overall token frequency of every obtained pair. The set of distributional contexts considered include the following:

- *Syn0*: this grouped together words which were not preceded by any element which prototypically introduces nouns in English (e.g., articles, possessive determiners, etc.). Thus, this category included nouns which occurred with no distributional context (or at least not a context that would assist word categorization) as well as verbs, adjectives, and adverbs. Words in the corpora whose distributional context matched this description scored 1 here and 0 in all the other contexts.
- *Syn1*: this grouped together words preceded by an indefinite article (e.g., *a dog*, *an apple*). Words within such a context scored 1 here and 0 elsewhere. Since it is local syntactic contexts that are being considered, an adjective such as *beautiful* in *a beautiful girl* would also score 1 here.
- *Syn2*: this was meant to group together words which were introduced by a definite article (e.g., *the toys*). Words found in such a syntactic context scored 1 here and 0 elsewhere.
- *Syn3*: words preceded by a demonstrative determiner were classified in this group (e.g., *this man*, *those birds*, etc.). Words within such a distributional context scored 1 here and 0 elsewhere.
- *Syn4*: this category included all words in the corpora which were preceded by any form of possessive determiner (i.e., *my hand*, *your teddy*, etc.). Words within such distributional contexts scored 1 in this category and 0 in all other categories.
- *Syn5*: this category was meant to group items which were introduced by any English quantifier (e.g., *many things*, *some cookies*, etc.). Items found within such contexts scored 1 in this category and 0 everywhere else.

## Results and Discussion

Parallel to the analysis with semantic cues in Study 1, for the analysis of distributional cues, the set of nouns and the set of non-nouns obtained from the corpora were tested using the distributional contexts described above. The total number of nouns that were found in each of the distributional contexts established is shown in **Table 5**, while **Table 6** shows the results obtained from the equivalent analysis with the rest of open class words.

As in Study 1, the categorization potential of the selected variables was assessed by means of a discriminant analysis. In order to examine the effects of the interaction between distributional and semantic cues, the set of six distributional

**TABLE 6 |** Total of non-nouns in each of the distributional contexts.

|                            | Total types | Total tokens | TTR   |
|----------------------------|-------------|--------------|-------|
| { $\emptyset$ } + x (Syn0) | 4,380       | 81,300       | 0.131 |
| {a, an} + x (Syn1)         | 346         | 2,009        | 0.172 |
| {the} + x (Syn2)           | 228         | 1,112        | 0.205 |
| {DEMONSTR.} + x (Syn3)     | 433         | 1,615        | 0.268 |
| {POSSESSIVE} + x (Syn4)    | 142         | 329          | 0.432 |
| {QUANTIFIER} + x (Syn5)    | 528         | 1,682        | 0.314 |

variables and the four semantic variables considered in this second study were introduced together as predictor variables in the discriminant function. Regarding types, when the combination of cues was introduced, there were 90.1% of correctly classified noun types and 96.8% of correctly classified non-noun types. Overall correct classification was 93.0% of all types, Wilks  $\lambda = 0.312$ ,  $\chi^2 = 11196.554$ ,  $p < 0.001$ .

The same classificatory system made up of the combination of distributional and semantic variables was tested with the set of tokens from the four child corpora. When all the variables were introduced simultaneously as predictor variables in a standard discriminant analysis, there were 47.9% of correctly classified noun tokens and 98.3% of correctly classified other open class word tokens. Overall correct classification reached 70.1%, Wilks  $\lambda = 0.899$ ,  $\chi^2 = 1028.161$ ,  $p < 0.001$ .

As can be seen, the results obtained with a combination of semantic and distributional cues are higher than the ones obtained with semantic cues only, both with types and tokens. Furthermore, higher scores of correct noun classification as well as correct non-noun classification are also obtained with the combination of distributional and semantic cues than with distributional cues alone (Feijoo et al., 2015).

In terms of accuracy (i.e., number of non-nouns which are correctly classified as such), the high results obtained in Study 1 are replicated here and they are not affected by the presence of distributional variables. In this way, we can claim that, on the basis of the information available in the input and the way grammatical categories are represented, children are very unlikely to misclassify verbs, adjectives or adverbs in the noun category. This would make a noun category very accurate and with a very low chance of including non-noun elements.

In terms of completeness, the results obtained in Study 2 indicate that children are more likely to create a more complete nominal category (i.e., one that includes many more nominal elements) when using distributional and semantic cues at the same time, rather than when using semantic cues alone. The analysis with types when using cues in combination reveals a very high proportion of correctly classified nouns (and, therefore, a very low risk of there being a misclassification of noun types in the non-noun group).

The analysis with tokens provides correct noun classification scores which are higher when using cues in combination rather than when using semantic cues alone. Even if such correct classification scores with noun tokens are still slightly below 50%, evidence from previous studies (Marchman and Bates, 1994; Bybee, 1995; Maratsos, 2000) suggests that regular

**TABLE 5 |** Total of nouns in each of the distributional contexts.

|                            | Total types | Total tokens | TTR   |
|----------------------------|-------------|--------------|-------|
| { $\emptyset$ } + x (Syn0) | 4,311       | 23,651       | 0.357 |
| {a, an} + x (Syn1)         | 1,459       | 6,414        | 0.227 |
| {the} + x (Syn2)           | 2,194       | 10,815       | 0.254 |
| {DEMONSTR.} + x (Syn3)     | 895         | 2,674        | 0.377 |
| {POSSESSIVE} + x (Syn4)    | 1,120       | 4,554        | 0.248 |
| {QUANTIFIER} + x (Syn5)    | 1,221       | 3,473        | 0.357 |

morphosyntactic patterns are generalized once patterns exhibit a relative type frequency. Thus, children appear to be more attentive to type frequency than to token frequency, and they are more likely to generalize from distributional contexts that appear on many stems than those that appear on only a few stems, even when the token instantiations of those fewer stems have an overall higher frequency. High token frequency is useful to keep an irregular form, but does not make a paradigm productive. On the other hand, type frequency helps language learners to identify productive paradigms (Clark, 2009; Ambridge and Lieven, 2011). All in all, we could claim that distributional and semantic cues available in child-directed speech interact in such a productive way as to allow the classification of most nouns in their correct grammatical category.

## GENERAL DISCUSSION

As seen earlier, the objective of the first study described in the present paper was to analyze the potential usefulness of semantic information as far as the categorization of English nouns is concerned. Traditional accounts on noun categorization based on semantic information have put forward the idea that young language learners might group all nouns together under the semantic label of “object” and all verbs together under the semantic label of “action” (Pinker, 1984). The fact that most nouns refer to common objects and their subsequent imageability based on notional grounds has been argued to be the reason why nouns are learned before verbs or before words encoding actions and relations in language development (Caselli et al., 1999; Bassano, 2000; Gleitman et al., 2005).

However, more recent findings show that nominal elements denoting common objects dominate children’s early vocabularies only as far as types are concerned (Nelson et al., 1993; Gopnik and Choi, 1995; Nelson, 1995). However, reports on early vocabulary production show that there are more tokens of non-nominal expressions (i.e., verbs and relational words such as *there*, *up* or *no*) than of nouns, and that this tendency is not only true of the English language (Gopnik and Choi, 1995).

Thus, as the authors suggest, the makeup of children’s early lexicons might not be the result of there being a more “learnable” or more “imaginable” category in semantic terms, but it might be a reflection of children’s actual linguistic experience, in such a way that children’s first words might be instantiations of the kind of words that their parents used with them. This is also coherent with the results obtained from the input analysis undertaken in the present studies. As seen earlier (see Table 2), the descriptive data that was obtained from the corpus preparation reveal that the kind of linguistic experience to which the four children under consideration are exposed contains far more nominal types than non-nominal types (i.e., there were overall 5,388 noun types and 4,233 other open class word types). However, when it comes to tokens, other open class words exceed nominal tokens by far (i.e., the four corpora together contained 51,577 nominal tokens, but 88,047 other open class word tokens).

The predominance of noun types over other open class word types has been explained by the fact that many very complex

and abstract entities are realized as nouns in adult language (Nelson, 1995). This raises the question of how many of these nominal types are actually “easily learnable.” If children engage in word categorization tasks guided by the fact that all words that refer to common objects are grouped together under the noun category, then how many of the overall 5,388 noun types considered in the present study can actually be described by these semantic features and how salient is that proportion in statistical terms?

Previous analyses of linguistic input addressed to young language learners have shown that common object nouns represent only a very small proportion of all the noun repertory that children hear (Nelson et al., 1993; Nelson, 1995; Monaghan and Mattock, 2012). The initial prediction as far as the first study was concerned was that many of the nouns to which very young English-learning children are exposed refer to basic-level common objects, and will be subsumed by variable *Sem2*. Other nouns were expected to either refer to proper names of people (i.e., *Sem1* nouns) or to non-discrete mass entities (i.e., *Sem3* nouns). Neither of them would pose any learning problems either. However, an important number of nominal elements were also predicted to lack any of the above-mentioned semantic characteristics (i.e., *Sem4* nouns). Without the necessary semantic components, those nouns would not guide children in their categorization tasks, provided children perform such tasks on the basis of semantic information.

These predictions are confirmed by the data obtained in the first analysis. The descriptive data from the first study show that only approximately half of the noun types as well as half of the nominal tokens refer to basic-level object nouns and were described by variable *Sem2*, while the other half belonged to the other three semantic subsets. Within those, about a third of the noun types and a quarter of the noun tokens were included in variable *Sem4*, which was the one which grouped together all nouns which did not have any of the semantic features that would foster correct grammatical categorization on the basis of semantic information alone. The results obtained from the discriminant analyses performed with types as well as with tokens confirm this weak correlation between semantic information and grammatical category assignment as far as nouns are concerned. Thus, for the type analysis, only 67.9% of the nominal types were correctly classified, and these completeness scores dropped dramatically in the token analysis, with only 39.9% of correctly classified nominal tokens.

A further objective in this first study was to see whether there was an overlap between nouns and other open class words as far as semantic information is concerned, and to test whether the overlap was significant enough so as to bring about a considerable misclassification of elements. Provided that, according to Pinker (1984), children engage in a semantic analysis of the input and make the hypothesis that all words that denote actions belong to the grammatical category of verbs, what do children do when they encounter action words which are not verbs? And what do they do with verbs that do not denote actions? Does the input offer a high proportion of contradictory information of this kind?

The present analysis shows that, indeed, there is a slight degree of semantic overlap between nouns and other open class words, since some of the XBLOC nouns have certain semantic features which are typical of verbs (i.e., mainly nouns that denote actions). At the same time, obviously not all of the other open class words denoted actions since, besides non-action verbs, adjectives, and adverbs were also included in this group. Thus, in the case of semantic information, the kind of overlap between nouns and other open class words seems to be one way, that is, some of the nouns might lack the corresponding nominal semantic features, or might have verbal semantic features, and might therefore be misclassified as other open class words (i.e., mainly misclassified as verbs). However, the same risk does not seem to hold for any of the other open class words, since they are very unlikely to contain any of the semantic features associated to nouns, and thus be misclassified as such on the basis of semantic information. Empirical analyses of children's early vocabularies also suggest that, even when the very same word can be both a noun and a verb (e.g., English *kiss*, *hug*, *call*, *help*) most children assign those action words exclusively to the verb category, regardless of their parents' frequent use of them as nouns (Nelson, 1995). As mentioned earlier, the discriminant analyses using semantic variables that were performed on all types and tokens confirm this, since the accuracy scores that were obtained in all cases were very close to 100% of correctly classified other open class words.

The fact that correct classification scores were far better with types than with tokens confirms the tendency described above in connection to early vocabulary production, where higher productivity of nominal types is observed, while tokens from other categories outnumber nominal tokens. Thus, children's early word production can be seen as a direct reflection of the kind of linguistic environment that they have experienced. However, nouns cannot be claimed to be "easily learnable" on the basis of their semantic association to basic-level objects alone, since statistical analyses where the diagnosticity of such semantic classification was tested provided a considerable proportion of misclassified nouns.

These findings suggest that semantic information may not always be the only factor which is used to determine the assignment of words to their grammatical category. Furthermore, as other studies have also suggested (Monaghan et al., 2007; Monaghan and Mattock, 2012; Yurovsky et al., 2012), learners never hear speech with just a single kind of cue to word categorization in isolation. In children's natural linguistic environment, there are multiple redundant language-specific cues to word category membership.

When distributional and semantic cues were put to interact with one another in the second study, overall successful categorization scores were expected to improve when compared to the results obtained with semantic cues in isolation. This prediction was born out by the results obtained from the second study. The interaction between semantic and distributional cues gave more successful results than semantic cues in isolation. The results obtained regarding successful categorization with both kinds of cues in combination were also higher than the results obtained in previous studies using distributional cues

alone (Mintz, 2003; Monaghan et al., 2007; Feijoo et al., 2015). In this sense, the results from both our first and our second study suggest that semantic cues contributed in providing an accurate grammatical category of nouns, since very few other open class words are at risk of being misclassified as nouns. On the other hand, distributional cues might have contributed in providing greater completeness scores, with a larger number of nouns being correctly classified as nouns, since the low completeness scores obtained from semantic cues alone were improved.

Monaghan et al. (2007) have already proposed the *Phonological-Distributional Coherence Hypothesis* in their analysis with distributional and phonological cues. According to them, both sources of information contribute differently toward word classification. Other studies have also highlighted the benefits of multiple types of cue for word categorization as well as for other language learning tasks (Monaghan et al., 2007; Smith et al., 2010; Riordan and Jones, 2011; Monaghan and Mattock, 2012; Yurovsky et al., 2012). Thus, having a combined number of different variables seems to increase the likelihood with which a given element will be successfully encoded and learned by children.

On the other hand, having to attend to several types of cues does not seem to imply an increase in terms of difficulty or cognitive processing demand on the part of very young language learners, at least as far as the combination of semantic and distributional information is concerned. The empirical evidence available to date seems to suggest that children are able to attend to multiple cues and use them for language learning tasks from a very early age (Thiessen and Saffran, 2003). In particular, by 14 months, infants have been described to be able to use determiners to identify nouns (Höhle et al., 2004; Shi and Melançon, 2010). Furthermore, hearing a word in a predictable distributional context promotes its identification for adults as well as for young infants (Lany, 2014). Thus, young language learners may also more readily encode a novel noun when it is preceded by a determiner. Facility with encoding would also make it easier for infants to determine the referent of the novel noun, and to form a robust mapping between the noun and the object it refers to, provided such mapping exists.

Therefore, on the basis of the evidence provided by the data obtained in our studies, we can claim that the combination of semantic and distributional information found in natural child-directed speech could significantly contribute to the correct categorization of most of the nouns to which English-learning infants are exposed. However, the results reported in our two studies are restricted to the kind of English speakers considered in our analyses. There might be important differences -both in terms of quantity and quality- in the input used with infants exposed to languages other than English, or infants from cultures other than western cultures and who are conventionally spoken to in a different way. Thus, for example, other languages with less distributional regularities might exploit phonological or prosodic information that might assist young language learners in their word categorization tasks. Further research on child-directed speech by non-English speakers should shed some light on how

the results obtained from the present studies would generalize cross-linguistically and cross-culturally.

## AUTHOR CONTRIBUTIONS

SF, CM, AA, and ES: Contributions to the conception and design; analysis and interpretation of data; drafting the work and revising it critically; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity are appropriately resolved.

## REFERENCES

- Ambridge, B., and Lieven, E. V. M. (2011). *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511975073
- Aslin, R. N., Woodward, J. Z., LaMendola, N. P., and Bever, T. G. (1996). "Models of word segmentation in fluent maternal speech to infants," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Mahwah, NJ: Erlbaum), 117–134.
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *J. Child Lang.* 20, 395–418. doi: 10.1017/S0305000900008345
- Bassano, D. (2000). Early development of nouns and verbs in French: exploring the interface between lexicon and grammar. *J. Child Lang.* 27, 521–559. doi: 10.1017/S0305000900004396
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., et al. (1994). Developmental and stylistic variation in the composition of early vocabulary. *J. Child Lang.* 21, 85–124. doi: 10.1017/S0305000900008680
- Benitez, V. L., Yurovsky, D., and Smith, L. B. (2016). Competition between multiple words for a referent in cross-situational word learning. *J. Mem. Lang.* 90, 31–48. doi: 10.1016/j.jml.2016.03.004
- Bernstein Ratner, N., and Rooney, B. (2001). "How accessible is the lexicon in Motherese?" in *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition*, eds J. Weissenborn and B. Höhle (Amsterdam: John Benjamins), 71–78. doi: 10.1075/lald.23.06ber
- Bybee, J. (1995). Regular morphology and the lexicon. *Lang. Cogn. Process.* 10, 425–455. doi: 10.1080/01690969508407111
- Cartwright, T. A., and Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition* 63, 121–170. doi: 10.1016/S0010-0277(96)00793-7
- Caselli, M. C., Casadio, P., and Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *J. Child Lang.* 26, 69–111. doi: 10.1017/S0305000998003687
- Clark, E. V. (2009). *First Language Acquisition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511806698
- Conwell, E., and Morgan, J. (2012). Is it a noun or is it a verb? Resolving the ambicategoricity problem. *Lang. Learn. Dev.* 8, 87–112. doi: 10.1080/15475441.2011.580236
- Feijoo, S., and Hilferty, J. (2013). The availability of cues for word segmentation and vocabulary acquisition in Catalan child-directed speech. *RLA* 51, 13–27. doi: 10.4067/S0718-48832013000200002
- Feijoo, S., Muñoz, C., and Serrat, E. (2015). Morphosyntactic cues to noun categorization in English child-directed speech. *Lang. Commun.* 45, 1–11. doi: 10.1016/j.langcom.2015.07.001
- Fitneva, S. A., Christiansen, M. H., and Monaghan, P. (2009). From sound to syntax: phonological constraints on children's lexical categorization of new words. *J. Child Lang.* 36, 967–997. doi: 10.1017/S0305000908009252
- Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press. doi: 10.1093/0198236360.001.0001
- Frank, M. C., Slemmer, J. A., Marcus, G. F., and Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Dev. Sci.* 12, 504–509. doi: 10.1111/j.1467-7687.2008.00794.x

## FUNDING

This work was supported by the University of Girona [grant number MPCUdG2016-123], and the Autonomous Catalan Government [grant number 2014SGR1089].

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01242/full#supplementary-material>

- Gelman, S. A., and Taylor, M. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child Dev.* 55, 1535–1540. doi: 10.2307/1130023
- Gerken, L. A., Wilson, R., and Lewis, W. (2005). Seventeen-month-olds can use distributional cues to form syntactic categories. *J. Child Lang.* 32, 249–268. doi: 10.1017/S0305000904006786
- Gleitman, L. (1990). The structural sources of verb meanings. *Lang. Acquis.* 1, 3–55. doi: 10.1027/s15327817la0101\_2
- Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Lang. Learn. Dev.* 1, 23–64. doi: 10.1207/s15473341ld0101\_4
- Gogate, L. J., Bahrick, L. E., and Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Dev.* 71, 878–894. doi: 10.1111/1467-8624.00197
- Gomez, R. L., and Lakusta, L. (2004). A first step in form-based category abstraction in 12-month-old infants. *Dev. Sci.* 7, 567–580. doi: 10.1111/j.1467-7687.2004.00381.x
- Gopnik, A., and Choi, S. (1995). "Names, relational words, and cognitive development in English and Korean speakers: nouns are not always learned before verbs," in *Beyond Names for Things: Young Children's Acquisition of Verbs*, eds M. Tomasello and W. E. Merriman (Mahwah, NJ: Lawrence Erlbaum Associates), 63–80.
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., and Schmitz, M. (2004). Functional elements in infants' speech processing: the role of determiners in the syntactic categorization of lexical elements. *Infancy* 5, 341–353. doi: 10.1207/s15327078in0503\_5
- Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., and Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *J. Child Lang.* 20, 523–549. doi: 10.1017/S0305000900008461
- Katz, N., Baker, E., and Macnamara, J. (1974). What's in a name? On the child's acquisition of proper and common names. *Child Dev.* 45, 269–273. doi: 10.2307/1127970
- Kelly, M. H. (1996). "The role of phonology in grammatical category assignments," in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, eds J. L. Morgan and K. Demuth (Mahwah, NJ: Erlbaum), 249–262.
- Lany, J. (2014). Judging words by their covers and the company they keep: probabilistic cues support word learning. *Child Dev.* 85, 1727–1739. doi: 10.1111/cdev.12199
- Lany, J., and Gomez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychol. Sci.* 19, 1247–1252. doi: 10.1111/j.1467-9280.2008.02233.x
- Lany, J., and Saffran, J. R. (2010). From statistics to meaning: infant acquisition of lexical categories. *Psychol. Sci.* 21, 284–291. doi: 10.1177/0956797609358570
- Laurence, S., and Margolis, E. (2001). The poverty of the stimulus argument. *Br. J. Philos. Sci.* 52, 217–276. doi: 10.1093/bjps/52.2.217
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*, 3rd Edn, Vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates.
- Maratsos, M. (1999). "Some aspects of innateness and complexity in language acquisition," in *The Development of Language*, ed. M. Barret (Hove: Psychology Press), 191–228.

- Maratsos, M. (2000). More overregularizations after all: new data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen, & Xu. *J. Child Lang.* 27, 183–212. doi: 10.1017/S0305000999004067
- Marchman, V., and Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *J. Child Lang.* 21, 339–366. doi: 10.1017/S0305000900009302
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90, 91–117. doi: 10.1016/S0010-0277(03)00140-9
- Monaghan, P., Chater, N., and Christiansen, M. H. (2005). The differential contribution of phonological and distributional cues in grammatical categorisation. *Cognition* 96, 143–182. doi: 10.1016/j.cognition.2004.09.001
- Monaghan, P., and Christiansen, M. H. (2010). Words in puddles of sound: modelling psycholinguistic effects in speech segmentation. *J. Child Lang.* 37, 545–564. doi: 10.1017/S0305000909990511
- Monaghan, P., Christiansen, M. H., and Chater, N. (2007). The phonological distributional coherence hypothesis: cross-linguistic evidence in language acquisition. *Cogn. Psychol.* 55, 259–305. doi: 10.1016/j.cogpsych.2006.12.001
- Monaghan, P., and Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition* 123, 133–143. doi: 10.1016/j.cognition.2011.12.010
- Nelson, K. (1995). “The dual category problem in the acquisition of action words,” in *Beyond Names for Things: Young Children’s Acquisition of Verbs*, eds M. Tomasello and W. E. Merriman (Mahwah, NJ: Lawrence Erlbaum Associates), 223–250.
- Nelson, K., Hampson, J., and Shaw, L. K. (1993). Nouns in early lexicons: evidence, explanations and implications. *J. Child Lang.* 20, 61–84. doi: 10.1017/S0305000900009120
- Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: MIT Press.
- Quine, W. (1960). *Word and Object*. Cambridge, MA: Harvard University Press.
- Redington, M., Chater, N., and Finch, S. (1998). Distributional information: a powerful cue for acquiring syntactic categories. *Cogn. Sci.* 22, 435–469. doi: 10.1207/s15516709cog2204\_2
- Riordan, B., and Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: comparing feature-based and distributional models of semantic representation. *Top. Cogn. Sci.* 3, 303–345. doi: 10.1111/j.1756-8765.2010.01111.x
- Rosch, E. (1978). “Principles of categorization,” in *Cognition and Categorization*, eds E. Rosch and B. B. Lloyd (Hillsdale, MI: Lawrence Erlbaum Associates), 27–48.
- Scott, R. M., and Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition* 122, 163–180. doi: 10.1016/j.cognition.2011.10.010
- Shi, R., and Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy* 15, 517–533. doi: 10.1111/j.1532-7078.2009.00022.x
- Smith, L. B., Colunga, E., and Yoshida, H. (2010). Knowledge as process: contextually-cued attention and early word learning. *Cogn. Sci.* 34, 1287–1314. doi: 10.1111/j.1551-6709.2010.01130.x
- Smith, L. B., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010
- Soja, N. N. (1992). Inferences about the meaning of nouns: the relationship between perception and syntax. *Cogn. Dev.* 7, 29–45. doi: 10.1016/0885-2014(92)90003-A
- Suanda, S. H., Mugwanya, N., and Namy, L. L. (2014). Cross-situational statistical word learning in young children. *J. Exp. Child Psychol.* 126, 395–411. doi: 10.1016/j.jecp.2014.06.003
- Tare, M., Shatz, M., and Gilbertson, L. (2008). Maternal uses of non-object terms in child-directed speech: color, number and time. *First Lang.* 28, 87–100. doi: 10.1177/0142723707085316
- Theakston, A. L., Lieven, E. V., Pine, J., and Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *J. Child Lang.* 28, 127–152. doi: 10.1017/S0305000900004608
- Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716. doi: 10.1037/0012-1649.39.4.706
- Tomasello, M. (2010). “Language development,” in *Blackwell Handbook of Childhood Cognitive Development*, ed. U. Goswami (Cambridge: Blackwell), 239–257.
- Vlach, H. A., and Johnson, S. P. (2013). Memory constraints on infants’ cross-situational statistical learning. *Cognition* 127, 375–382. doi: 10.1016/j.cognition.2013.02.015
- Yu, C. (2006). “Learning syntax–semantics mappings to bootstrap word learning,” in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, ed. R. Sun (Mahwah, NJ: Lawrence Erlbaum Associates), 924–929.
- Yu, C., and Ballard, D. H. (2007). A unified model of early word learning: integrating statistical and social cues. *Neurocomputing* 70, 2149–2165. doi: 10.1016/j.neucom.2006.01.034
- Yurovsky, D., Yu, C., and Smith, L. B. (2012). Statistical speech segmentation and word learning in parallel: scaffolding from child-directed speech. *Front. Psychol.* 3:374. doi: 10.3389/fpsyg.2012.00374

**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 © 2017 Feijoo, Muñoz, Amadó and Serrat. 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.



# Language Proficiency and Sustained Attention in Monolingual and Bilingual Children with and without Language Impairment

Tessel Boerma<sup>1\*</sup>, Paul Leseman<sup>1</sup>, Frank Wijnen<sup>2</sup> and Elma Blom<sup>1</sup>

<sup>1</sup> Department of Special Education, Utrecht University, Utrecht, Netherlands, <sup>2</sup> Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands

## OPEN ACCESS

### Edited by:

Theo Marinis,  
University of Reading,  
United Kingdom

### Reviewed by:

Arturo Hernandez,  
University of Houston, United States  
Marie Lallier,  
Basque Center on Cognition, Brain  
and Language, Spain

### \*Correspondence:

Tessel Boerma  
t.d.boerma@uu.nl

### Specialty section:

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

**Received:** 24 February 2017

**Accepted:** 06 July 2017

**Published:** 21 July 2017

### Citation:

Boerma T, Leseman P, Wijnen F and  
Blom E (2017) Language Proficiency  
and Sustained Attention  
in Monolingual and Bilingual Children  
with and without Language  
Impairment. *Front. Psychol.* 8:1241.  
doi: 10.3389/fpsyg.2017.01241

**Background:** The language profiles of children with language impairment (LI) and bilingual children can show partial, and possibly temporary, overlap. The current study examined the persistence of this overlap over time. Furthermore, we aimed to better understand why the language profiles of these two groups show resemblance, testing the hypothesis that the language difficulties of children with LI reflect a weakened ability to maintain attention to the stream of linguistic information. Consequent incomplete processing of language input may lead to delays that are similar to those originating from reductions in input frequency.

**Methods:** Monolingual and bilingual children with and without LI ( $N = 128$ ), aged 5–8 years old, participated in this study. Dutch receptive vocabulary and grammatical morphology were assessed at three waves. In addition, auditory and visual sustained attention were tested at wave 1. Mediation analyses were performed to examine relationships between LI, sustained attention, and language skills.

**Results:** Children with LI and bilingual children were outperformed by their typically developing (TD) and monolingual peers, respectively, on vocabulary and morphology at all three waves. The vocabulary difference between monolinguals and bilinguals decreased over time. In addition, children with LI had weaker auditory and visual sustained attention skills relative to TD children, while no differences between monolinguals and bilinguals emerged. Auditory sustained attention mediated the effect of LI on vocabulary and morphology in both the monolingual and bilingual groups of children. Visual sustained attention only acted as a mediator in the bilingual group.

**Conclusion:** The findings from the present study indicate that the overlap between the language profiles of children with LI and bilingual children is particularly large for vocabulary in early (pre)school years and reduces over time. Results furthermore suggest that the overlap may be explained by the weakened ability of children with LI to sustain their attention to auditory stimuli, interfering with how well incoming language is processed.

**Keywords:** language impairment, bilingualism, sustained attention, vocabulary, morphology

## INTRODUCTION

There is enormous variation in children's rates and courses of language development, caused by the interplay of child-internal factors with a genetic basis (Stromswold, 2001), and child-external factors in the environment (Hoff, 2006). Child-internal and child-external factors can influence language outcomes in comparable ways, as is illustrated by the partially overlapping language profiles of children with an inborn primary or specific language impairment (further on called LI) and children who are raised bilingually. Profound language delays have been documented for both children with LI (Rice, 2004; Krok and Leonard, 2015; Rice and Hoffman, 2015) and bilingual children (Bialystok et al., 2010; Farnia and Geva, 2011; Paradis et al., 2016), and comparisons of these two groups of children showed strikingly similar performance on core language domains, such as vocabulary and morphology (Grüter, 2005; Paradis, 2005; Blom and Boerma, 2017). It is, however, unknown whether these similarities are temporary and limited to certain developmental stages. The first aim of the present study was therefore to compare the effects of LI and bilingualism on children's language skills over time.

The second aim of the current study was to better understand why the language profiles of children with LI and bilingual children show overlap, so as to shed light on the underlying causes of the effects of LI on children's language proficiency. Although the origins of the language delays are evidently different for the two groups of children, language input may play a key role in both. The quantity of language input is one of the most important factors contributing to the acquisition of language (Hart and Risley, 1995; Hoff, 2006) and it is well-established that the language outcomes of bilingual children are affected by the distributed nature of their input over two (or more) languages (e.g., Hoff et al., 2012). The language skills of children with LI may be poor due to an impaired capacity to process language input efficiently (e.g., Leonard et al., 2007b). Deficits in domain-general cognitive mechanisms are thought to underlie this limited input processing capacity, and particularly working memory has been frequently associated with the language difficulties of children with LI (for reviews, see Montgomery et al., 2010; Henry and Botting, 2016). There are furthermore intuitive and empirical reasons to assume interaction between language acquisition and attention mechanisms (Yoshida et al., 2011; Kapa and Colombo, 2014), which are tightly connected to working memory (Cowan, 1995; Baddeley, 2000), but less is yet known about this relation in children with LI.

A conceivable hypothesis is that the language problems of children with LI reflect a weakened ability to maintain attention to the stream of linguistic information, leading to incomplete processing of language input. In light of findings showing that children with LI have poor sustained attention (Ebert and Kohnert, 2011), the current study tested this hypothesis within a monolingual and bilingual context. We investigated the effects of LI and bilingualism on children's auditory and visual sustained attention skills, and explored the role of sustained attention in explaining the effects of LI on children's language outcomes. We hereby aimed to elucidate the relation between

the linguistic and non-linguistic deficits of children with LI, which is a necessary step in further understanding the nature of the disorder (Kapa and Plante, 2015). Below, we first review research on the language development of bilingual children and children with LI, and discuss possible origins of their language delays. Subsequently, the relation between language and sustained attention is addressed. Throughout, we focus on the domains of vocabulary and morphology, as these are both considerably affected by LI and reduced input due to bilingualism (e.g., Blom and Boerma, 2017), and are subject of investigation in the present research.

It is well-documented that children who learn two or more languages, either from birth or later in childhood, lag behind their monolingual peers when only one of their languages is evaluated (Thordardottir et al., 2006; Scheele et al., 2010; Hoff et al., 2012). In early stages of acquisition, bilingual toddlers show slower rates of language-specific growth than monolingual toddlers, particularly in the domain of vocabulary which has been studied most often (Vagh et al., 2009; Silvén et al., 2014), but also in terms of grammar knowledge (Hoff et al., 2012). The consequent delays appear persistent, as is demonstrated by longitudinal research with bilingual (pre)schoolers (Farnia and Geva, 2011; Paradis et al., 2016). Tracking children's vocabulary growth in English from grade 1 to 6, Farnia and Geva (2011) observed that their bilingual participants who learned English as a second language did not fully catch up with the monolingual controls, even though the bilinguals had a steeper learning curve in the primary grades and thus seemed to benefit from the increasing exposure to English at school. These findings correspond to results from other studies which indicate persistent gaps between the language-specific vocabulary size of monolingual and bilingual children (Appel and Vermeer, 1998; Cobo-Lewis et al., 2002; Roessingh and Elgie, 2009; Bialystok et al., 2010; Scheele, 2010; Thordardottir and Juliusdottir, 2013).

With respect to morphology, Paradis et al. (2016) also showed large and consistent delays over time, comparing bilingual children with monolingual norms. Around 60% of their Chinese-English participants did not achieve monolingual-like performance on an English verb morphology task after 6½ years of English schooling (see also, Jia and Fuse, 2007), and, in addition, growth curves suggested plateau effects. Low saliency in the input may render verb morphology notoriously difficult for children learning English as a second language, as Paradis et al. (2016) suggest. Moreover, English verb inflection can be extra challenging for children who cannot benefit from the presence of tense and agreement morphology in their first language, like children with a Chinese background (Paradis, 2011; Blom et al., 2012). Using the same participant sample as Paradis et al. (2016) but including more general standardized measures of English vocabulary and grammar knowledge, Paradis and Jia (2017) reported monolingual-like attainment for the majority of children on the majority of measures after 5½ years of English schooling. The persistence of bilingual children's language delays may thus, next to language background, depend on linguistic subdomain.

Paradis and Jia (2017) furthermore found that children's language environment, including amount and richness of English

input, predicted their language abilities and convergence to monolingual norms. These findings connect to a multitude of studies which established that the amount and quality of language-specific input is a strong determinant of skills in that language (see, Hart and Risley, 1995; Huttenlocher et al., 2002; Rowe, 2012; Grüter and Paradis, 2014), and the distributed nature of bilingual children's input is thereby one of the most important explanations for their documented language delays (Scheele et al., 2010; Hoff et al., 2012). Children's scores on measures of vocabulary (Scheele et al., 2010; Chondrogianni and Marinis, 2011; Hoff et al., 2012) and morphology (Paradis, 2010a; Blom et al., 2012; Thomas et al., 2014) have both been related to amount of exposure, but there are indications that certain morphological structures are less susceptible to input effects than vocabulary (Chondrogianni and Marinis, 2011). Lexical items need to be learned one-by-one and can thus only be successfully acquired through repeated exposure to the same form. In contrast, (regular) morphology is largely based on rule learning and allows for fast generalization to new forms. This makes morphology possibly less sensitive to limited exposure, and thus bilingualism, than vocabulary, although bilingual performance also highly depends on other factors, such as the frequency and complexity of linguistic structures (Paradis, 2010a; Rispen and de Bree, 2015). In particular, structures that are low in frequency and high in complexity may be strongly influenced by reduced input.

An inborn LI disproportionately affects a child's ability to learn language, in the absence of any clearly discernable cause (Leonard, 2014). Vocabulary is one domain in which delays are found (Rice and Hoffman, 2015), but LI is often more strongly associated with severe grammar weaknesses, especially in the domain of morphology (e.g., Rice et al., 1998; Ullman and Pierpont, 2005). Longitudinal work by Rice (2004, 2012) and Rice and Hoffman (2015) indicates that the delayed onset of language, characteristic of children with LI, is typically larger for grammar than for vocabulary. Once underway, both the lexical and grammatical development of children with LI seem to parallel the development of typically developing (TD) children.

Rice and Hoffman (2015) modeled the growth trajectories of children's receptive vocabulary over nearly two decades. A consistently lower level of performance for the children with LI in comparison with their TD peers was found, but both groups had a generally similar growth curve. Only in the pre-adolescent period, rate of acquisition decelerated in children with LI. Similar growth patterns for children with TD and LI were also reported for measures of grammatical development, including the production and grammatical judgment of finiteness markings (Rice, 2012). The children with LI eventually reached, much later than TD peers, adult-like ceiling performance for production, but the more difficult judgment task remained problematic into adolescence. These findings from research by Rice and colleagues are in agreement with other large-scale longitudinal work with children with LI which showed persistent language delays and stability of growth in this population (Beitchman et al., 1996; Johnson et al., 1999), with differences in initial severity determining long-term language outcomes (Law et al.,

2008; Conti-Ramsden et al., 2012). Moreover, these findings also correspond to recent work by Paradis et al. (2017) who compared the acquisition of tense morphology over time by bilingual children with and without LI, indicating developmental trajectories parallel to monolinguals with and without LI.

Several theories have been postulated to explain these persistent language delays of children with LI (see, Leonard, 2014). The current study, aiming to better understand the overlap between the language profiles of children with LI and bilingual children, will focus on accounts of LI that view the disorder as a problem of input or information processing (Kail, 1994; Leonard et al., 1997, 2007b). While factors in a child's social context, like bilingualism, produce variation in the language input of a child, in turn influencing the child's language development, it may be that an inborn LI leads to differences in how children can make use of the input. This hypothesis is based on findings from a growing body of work which suggests that problems of children with LI extend beyond linguistic domains (e.g., Henry et al., 2012; Vissers et al., 2015). Studies within the limited processing capacity framework have tried to integrate the linguistic and non-linguistic weaknesses of children with LI. Deficits in cognitive and perceptual mechanisms that are important for the acquisition of language, such as memory (Gathercole, 2006; Leonard et al., 2007b; Montgomery et al., 2010; Conti-Ramsden et al., 2015), and/or general speed of processing (Miller et al., 2001; Leonard et al., 2007b), may lead to incomplete or inadequate processing of the input, resulting in persistent language delays. As "cases of incomplete processing are assumed to be the functional equivalent of reductions in input frequency" (Leonard, 2014; p. 289), children with LI would need more exposure than their TD peers to successfully acquire language. This hypothesis is confirmed by several studies within the context of word learning (Rice et al., 1994; Gray, 2003; Riches et al., 2005; for a meta-analysis, see Kan and Windsor, 2010), and is furthermore supported by research on grammar acquisition showing that the effect of LI is more pronounced on low frequency than high frequency structures (Leonard et al., 2007a; Leroy et al., 2013).

A number of studies investigated the implications of these input dependencies for the language outcomes of bilingual children with LI, who are assumed to have a weaker capacity to process input efficiently compared with TD children, in addition to receiving less exposure in each language compared with monolingual children. Research conducted in the Netherlands showed that bilingual children with LI performed weaker on Dutch vocabulary and morphology tasks relative to both bilingual TD children and monolingual children with LI, indicating double delays (Verhoeven et al., 2011; Blom and Boerma, 2017). While the effect of LI on vocabulary scores was even larger in a bilingual than in a monolingual group of children, difficulty with morphology was not aggravated by the presence of LI in combination with bilingualism (Blom and Boerma, 2017; see also, Paradis, 2010b). Together with work that did not identify a double delay of bilingual children with LI on morphology (Paradis, 2007; Gutiérrez-Clellen et al., 2008; Rothweiler et al., 2012; Blom et al., 2013; Paradis et al., 2017), this supports the possibility that morphology is less susceptible to input effects than vocabulary (Chondrogianni and Marinis, 2011). However, the

mixed findings within the domain of morphology also indicate that input effects may not always function linearly (Conti-Ramsden, 2010) and, in addition, that other factors are likely to play a role in explaining the performance patterns of bilingual children with LI, including the type of target structure and the characteristics of the bilingual sample (Gathercole, 2010; Paradis, 2010b).

Within the limited input processing capacity framework, working memory has been most frequently studied to account for the language difficulties of children with LI. There is substantial evidence for working memory problems in children with LI and several studies have found associations between working memory and language, pointing to a possible and plausible cause of the weakened language skills of these children (for a recent review, see Henry and Botting, 2016). Next to working memory, the role of attention resources in children with LI is a focus of recent research. Attention is a basic cognitive capacity which is difficult to reduce to a single definition. It can refer to a person's ability to be alert, maintain focus over time, and selectively process relevant stimuli (Gomes et al., 2000). Common conceptualizations of attention imply strong connections between attention and language learning (for a review, see Ebert and Kohnert, 2011). For example, attention may be needed to direct a learner's focus to relevant linguistic stimuli in the input before they can be processed, and to maintain this focus in order to prevent reduced or incomplete processing of that input. Moreover, it has been hypothesized that the ability to engage and disengage attention at a fast pace is necessary for the processing of rapidly presented stimulus sequences (Hari and Renvall, 2001), which is characteristic of language input. Empirical support for the role of attention in language learning has been provided by several studies, associating attention mechanisms with artificial word learning (Yoshida et al., 2011; Kapa and Colombo, 2014) and speech processing (see, Stevens and Bavelier, 2012) in TD children. Together with the high comorbidity rate between children with LI and children with attention deficits (Tirosh and Cohen, 1998), this explains the interest to attention in the LI literature.

A growing body of work suggests that, next to having working memory deficits, children with LI also have a limited attention capacity compared with their TD peers, even in children without comorbid attention deficit (hyperactivity) disorder (Marton, 2008; Ebert and Kohnert, 2011). Children with LI have particularly often been found to perform poorly on tasks tapping into sustained attention (for a meta-analysis, see Ebert and Kohnert, 2011). There is strong evidence that children with LI have a weak ability to maintain their focus on auditory stimuli during a prolonged period of time (Noterdaeme et al., 2001; Dodwell and Bavin, 2008; Spaulding et al., 2008). In addition, problems with visual sustained attention have also been reported (Finneran et al., 2009), although the effects of LI are smaller in comparison with the auditory domain and findings are mixed (Ebert and Kohnert, 2011).

A number of studies also examined the relationship between the poor language and sustained attention skills of children with LI, finding positive associations. Work by Montgomery showed

that auditory sustained attention accounted for more than 45% of the variance in the online sentence processing of children with LI (Montgomery, 2008), and correlated highly with simple and complex sentence comprehension (Montgomery et al., 2009). Moreover, both auditory and visual sustained attention were positively correlated with picture-naming performance of children with LI and TD (Jongman et al., 2016), and auditory sustained attention was furthermore found to be associated with story generation skills (Duinmeijer et al., 2012). Blom and Boerma (2016) also investigated narrative abilities and showed that the effect of LI on story generation was mediated by sustained attention, measured with an integrated auditory and visual continuous performance task (CPT). Finally, findings from two intervention studies by Ebert et al. (2012, 2014) suggest that a treatment program designed to improve the processing speed and sustained attention skills of children with LI positively influenced children's language scores. These studies thus support the possibility that the language delays of children with LI reflect, at least in part, a weakened ability to maintain attention to the stream of linguistic information, interfering with how well language input is processed. The present study will extend this research and investigate the role of auditory and visual sustained attention in explaining the effect of LI on two core language domains, vocabulary and morphology, which are known to be affected by LI and by reduced input.

The current study will analyze this within both a monolingual and bilingual context. As of yet, few studies have examined sustained attention in bilingual children. Although bilingual children have been reported to outperform their monolingual peers on different attention tests, especially those involving conflict processing (e.g., Bialystok, 1999; Engel de Abreu et al., 2012), the so-called bilingual advantage is not ubiquitous (e.g., Duñabeitia et al., 2014) nor undisputed (Paap et al., 2015). A specific bilingual benefit on sustained attention in children has not yet been attested and the few adult studies reveal mixed findings (Bialystok et al., 2008; Krizman et al., 2012; Bak et al., 2014), emphasizing the need for further research. In addition, work on the relation between sustained attention and language in bilingual children with LI is sparse, only including the intervention studies of Ebert et al. (2012, 2014) with Spanish-English bilingual participants with LI. Like the work with monolingual samples (Montgomery, 2008; Montgomery et al., 2009; Duinmeijer et al., 2012; Blom and Boerma, 2016; Jongman et al., 2016), these studies suggest that sustained attention may also contribute to the language difficulties of children with LI growing up in bilingual learning settings. The current research will further explore this.

The first aim of the present study was to investigate whether the overlap between the language profiles of children with LI and bilingual children was temporary, or persisted over time. We used a four-group design, including monolingual and bilingual children with and without LI, which allowed for a systematic examination of the effects of LI and bilingualism on children's language development. We focused on children's vocabulary and morphology outcomes. Negative effects of LI were expected to emerge on both language domains (Krok and Leonard, 2015;

Rice and Hoffman, 2015), although larger effects were anticipated on morphology (Rice, 2012). Given the young age of the participants (5–8 years old) and the relatively short time span of the current study (3 years), effects of LI were furthermore assumed to remain stable over time (Rice, 2012). Vocabulary and morphology were also predicted to be negatively affected by reductions in input frequency as a result of bilingualism (Hoff et al., 2012; Paradis et al., 2016), with possibly more pronounced effects on vocabulary than morphology (Chondrogianni and Marinis, 2011). The gap between the monolinguals and bilinguals was not expected to fully close within the time frame of this study, but the effect of bilingualism may diminish over time due to accumulating input in school (Farnia and Geva, 2011).

The second aim of the current study was to better understand why similarities between the language profiles of children with LI and bilingual children emerge. We tested the hypothesis that the language difficulties of children with LI stem from auditory sustained attention deficits, since consequent incomplete processing of language input may lead to delays that are comparable to those originating from reductions in input frequency due to bilingualism. Visual sustained attention was also assessed to examine possible domain-general origins. Furthermore, the hypothesis was tested within a monolingual and bilingual context. The presence of LI was predicted to impact children's sustained attention skills, with relatively better performance of children with LI on the visual compared with the auditory domain (Ebert and Kohnert, 2011). Sustained attention was not hypothesized to be strongly influenced by bilingualism, although positive effects were considered possible in view of the literature on the cognitive benefit of bilingualism (e.g., Bialystok, 1999).

Previous work with children with LI showed that limitations in sustained attention are predictive of narrative skills (Blom and Boerma, 2016), and associated with sentence processing (Montgomery, 2008) and picture-naming (Jongman et al., 2016). We anticipated that sustained attention, and in particular auditory sustained attention, would also play a role in explaining the effect of LI on two core language areas, i.e., vocabulary and morphology, which are known to be influenced by a limited amount of input (e.g., Scheele et al., 2010; Blom et al., 2012) and thus likewise by the functional equivalent: incomplete processing of input. Given our hypothesis that the language delays of children with LI arise from a weakened ability to maintain attention to the stream of linguistic information, interfering with efficient input processing, effects of visual sustained attention were expected to be limited. Moreover, the impact of sustained attention deficits on morphology could be less pronounced in comparison with vocabulary, as previous work showed that morphology is less susceptible to input effects than vocabulary (Chondrogianni and Marinis, 2011). However, this may also depend on the frequency and complexity of the targeted structures (Paradis, 2010a; Rispens and de Bree, 2015). Finally, we had no clear theoretical or empirical reasons to assume substantial differences between the role of sustained attention in explaining the effect of LI on monolingual or bilingual children's language skills.

## MATERIALS AND METHODS

### Participants

The data from the current study were collected within a large-scale longitudinal project that aimed to investigate the linguistic and cognitive development of children with diverse language backgrounds in the Netherlands. Four groups, monolingual and bilingual children with and without LI, were followed from 2014 to 2016 and tested once a year (mean = 11 months). Children were around age 5 or 6 at the first wave of testing, and around age 7 or 8 at the third and last wave. For the present study, a matched subsample of this large-scale project was selected to be able to control for factors such as age, non-verbal intelligence (NVIQ) and socio-economic status (SES) when comparing different groups of children, as these factors may influence children's language skills (Hart and Risley, 1995; Conti-Ramsden et al., 2012). The group of bilingual children with LI (BILI) was the smallest ( $N = 33$ ) and therefore the basis for our participant match. Before wave 3, one child in the BILI group transferred to a school for children with an intellectual disability and was therefore excluded from the current study, resulting in groups of 32 children each (total  $N = 128$ ). Each child in the BILI group was matched on age in months at wave 1 to a bilingual typically developing child (BITD), a monolingual typically developing child (MOTD), and a monolingual child with LI (MOLI). As the BILI group had a relatively large age range and was on average slightly older than the other groups, it was not possible to find a close age match (i.e., a difference of less than 4 months) for all children. Some children were therefore matched on group level, aiming to form groups that were on average as comparable as possible. To this end, groups were furthermore matched on (in order of priority) NVIQ, exposure to Dutch (for the bilinguals), SES, and gender.

Group characteristics are displayed in **Table 1**. There were no significant differences between the four groups of children in age in months at wave 1 [ $F(3,124) = 0.25$ ,  $p = 0.86$ ,  $\eta_p^2 < 0.01$ ], wave 2 [ $F(3,124) = 0.03$ ,  $p = 0.99$ ,  $\eta_p^2 < 0.01$ ], nor wave 3 [ $F(3,124) = 0.07$ ,  $p = 0.98$ ,  $\eta_p^2 < 0.01$ ]. NVIQ, measured with the short version of the *Wechsler Nonverbal-NL* (Wechsler and Naglieri, 2008), did not significantly differ between the groups of children either [ $F(3,124) = 1.02$ ,  $p = 0.39$ ,  $\eta_p^2 = 0.02$ ]. In addition, no differences emerged in SES [ $H(3) = 5.5$ ,  $p = 0.14$ ], which was indexed by the average education level of the child's parents, measured on a nine-point scale (ranging from 1 'no education' to 9 'university degree'). There were also no gender differences between the four groups of children [ $\chi^2(3, N = 128) = 6.4$ ,  $p = 0.09$ ], although there was a relatively large number of boys in the groups of children with LI. Finally, the bilingual groups did not significantly differ in exposure to Dutch before the age of 4 [ $F(1,61) = 0.68$ ,  $p = 0.41$ ,  $\eta_p^2 = 0.01$ ], nor current exposure to Dutch at home [ $F(1,62) = 2.5$ ,  $p = 0.12$ ,  $\eta_p^2 = 0.04$ ]. The *Questionnaire for Parents of Bilingual Children* (PaBiQ; Tuller, 2015), administered at wave 1, measured the exposure to Dutch before the age of 4 as the percentage of input in Dutch that the child received before this age (both inside and outside home context), relative to the total amount of language

**TABLE 1 |** Demographic characteristics of the participants.

|      | N  | Age (months)<br>Wave 1 |  | Age (months)<br>Wave 2 |  | Age (months)<br>Wave 3 |  | Non-verbal IQ<br>(standardized) | Socio-economic<br>Status |           | Gender      |            | Exposure to Dutch<br>before age 4 (in %) | Exposure to Dutch at<br>Wave 1 (in %) |
|------|----|------------------------|--|------------------------|--|------------------------|--|---------------------------------|--------------------------|-----------|-------------|------------|------------------------------------------|---------------------------------------|
|      |    | Mean (SD)              |  | Mean (SD)              |  | Mean (SD)              |  |                                 | Mean (SD)                |           | N_GIRLS (%) | N_BOYS (%) |                                          |                                       |
| MOTD | 32 | 70.9 (7.0)             |  | 82.5 (6.9)             |  | 94.1 (6.9)             |  | 100.4 (11.9)                    |                          | 6.3 (2.0) |             | 14 (44%)   | 18 (56%)                                 | n/a                                   |
| MOLI | 32 | 71.4 (6.3)             |  | 82.8 (6.5)             |  | 94.6 (6.6)             |  | 97.5 (12.9)                     |                          | 5.2 (1.8) |             | 8 (25%)    | 24 (75%)                                 | n/a                                   |
| BITD | 32 | 71.3 (7.3)             |  | 83.0 (7.1)             |  | 94.8 (7.1)             |  | 95.8 (15.0)                     |                          | 5.3 (2.3) |             | 17 (53%)   | 15 (47%)                                 | 50.9 (12.0)                           |
| BILI | 32 | 72.4 (8.6)             |  | 83.0 (8.9)             |  | 94.7 (8.8)             |  | 94.7 (15.3)                     |                          | 5.7 (2.2) |             | 10 (31%)   | 22 (69%)                                 | 45.2 (16.5)                           |

MOTD, monolingual typically developing; MOLI, monolingual language impaired; BITD, bilingual typically developing; BILI, bilingual language impaired.

input. The PaBiQ measured current exposure to Dutch at home as the percentage of input in Dutch, relative to the total amount of language input, that the child heard from its mother, father, siblings, and other adults that had frequent contact with the child.

### Criteria for LI

All children in the MOLI and BILI groups had been diagnosed with LI before the start of this research. They were diagnosed with LI by licensed clinicians according to the standardized criteria that are used in the Netherlands. In the Netherlands, a child officially meets the criteria for LI when (s)he obtains a score of at least 2 standard deviations (SD) below the mean on an overall score of a standardized language assessment test battery or a score of at least 1.5 SD below the mean on two out of four subscales of this standardized language assessment (Stichting Siméa, 2014). The most commonly used test batteries include the Dutch version of the *Clinical Evaluation of Language Fundamentals* (CELF-4-NL; Kort et al., 2008), the *Schlichting Test for Language Production and Comprehension* (Schlichting and Lutje Spelberg, 2010a,b), and the *Dutch Language Proficiency Test for All Children* which has bilingual norms [Taaltoets Alle Kinderen (TAK); Verhoeven and Vermeer, 2001]. In addition, a guideline focusing on the assessment of bilingual children is provided by Stichting Siméa (2016), stating the need for a bilingual anamnesis and, if possible, evaluation of the first and second language.

At wave 1 and 2, all 64 children in the MOLI and BILI groups met the criteria for LI that were specified above. At wave 3, eight children (four bilingual and four matched monolingual children) did not meet these criteria anymore, confirming the fluid developmental pathways for language (Reilly et al., 2014). Given their history of LI and the long-term persistence of the language problems (Scarborough and Dobrich, 1990), we did not exclude these children. All children who participated in the present study had no intellectual disability (NVIQ range from 70 to 130), hearing impairment, severe articulatory difficulties or diagnosed attention deficit disorder. At the start of the research, 63 children with LI attended special education and one child with LI attended regular education with ambulatory care. During the study, 14 children with LI (five bilingual and nine monolingual) transferred from special to regular education. All TD children attended regular elementary schools and did not have documented language problems.

### Criteria for Bilingualism

Information about the home language environment of the children was provided by the parental questionnaire (PaBiQ; Tuller, 2015). A child was assigned to the monolingual group if both parents were native speakers of Dutch and always spoke Dutch to the child. A child was considered bilingual if at least one parent was a native speaker of another language than Dutch and spoke their mother tongue with the child for an extensive period of the child's life. All bilingual children who participated in this study were born in the Netherlands and learned Dutch as a second language. As elementary school starts at age 4 in the Netherlands, all children had received at least approximately 1 year of schooling in Dutch before the first wave of testing. The first languages of the bilingual TD children included Turkish

( $N = 14$ ), Tarifit-Berber ( $N = 10$ ), and Moroccan Arabic ( $N = 8$ ). The first languages of the bilingual children with LI were Turkish ( $N = 10$ ), Moroccan Arabic ( $N = 7$ ), Egyptian Arabic ( $N = 3$ ), Tarifit-Berber ( $N = 2$ ), Dari ( $N = 2$ ), Chinese ( $N = 1$ ), Pashto ( $N = 1$ ), Suryoyo ( $N = 1$ ), Kirundi ( $N = 1$ ), Russian ( $N = 1$ ), Portuguese ( $N = 1$ ), Danish ( $N = 1$ ), and Frisian ( $N = 1$ ).

## Materials and Procedures

The current study was part of a large-scale project which was approved by the Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University. Parents of participants signed an informed consent form. Children were individually tested in a quiet room at their school. Trained research assistants followed a strict protocol and administered a test battery, consisting of language, memory and attention tasks, in two separate sessions. Each test session lasted approximately 1 h. Receptive vocabulary, morphology and sustained attention were all assessed in the second session. Similar procedures were used at each wave of testing.

### Language

Receptive vocabulary was tested at all three waves with the *Peabody Picture Vocabulary Test* (PPVT-III-NL; Schlichting, 2005), which is a standardized test designed for a wide age range (2;3–90 years). Participants hear a target word and have to pick the correct referent out of four pictures. The task is divided in 17 sets, which increase in difficulty, with 12 target words in each set. We administered the PPVT-III-NL according to the official guidelines and thus determined the starting set based on a child's age. The task was terminated when a child picked the incorrect referent picture nine or more times in a set. Raw scores were used in the analyses.

Grammatical morphology was assessed at all three waves with a subtest of the *Dutch Language Proficiency Test for All Children* (TAK; Verhoeven and Vermeer, 2001), suitable for children aged 4–9. The subtest 'Word Formation' elicits 12 noun plurals and 12 past participles, including both regularly and irregularly inflected nouns and verbs. Children are presented with a picture and asked to finish an incomplete sentence uttered by the experimenter, hereby eliciting the plural of a noun (e.g., *Dit is één lepel, dit zijn twee...? Lepels*. [This is one spoon, these are two...? Spoons]) or the past participle of a verb (e.g., *Hier zie je Paul op de bank zitten. Gisteren heeft hij ook al op de bank...? Gezeten*. [Here you see Paul sitting on the couch, yesterday he has also... on the couch? Sat.]). Accuracy was scored offline by a native speaker of Dutch and the number of items correct (maximum = 24) was used in the analyses.

### Sustained Attention

Sustained attention was measured at wave 1 with an integrated visual and auditory CPT, which was based on the IVA+Plus (Sandford and Turner, 2004) and identical to the task used in Blom and Boerma (2016). The task was administered on a laptop using the experimental software E-Prime 2.0 (Schneider et al., 2002). Children were presented with visual and auditory stimuli that could either be a target (number '1') or a distractor (number '2'). Each visual stimulus was presented for 167 milliseconds.

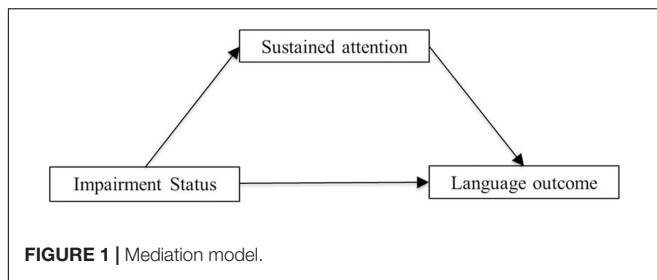
Irrespective of modality, children were asked to press the space bar in response to a target stimulus, but to refrain from responding when a distractor appeared. The test included 168 trials, excluding the practice phase, in which visual and auditory targets ( $N = 84$ ) and distractors ( $N = 84$ ) were mixed and presented randomly. The task lasted approximately 10 min, during which children were required to stay alert and maintain their attention.

Response sensitivity on this task was scored as  $d'$  (Macmillan and Creelman, 2005). For visual sustained attention, this inherently dual score reflects percent correct responses to visual targets (hits) relative to percent incorrect responses to visual distractors (false alarms). For auditory sustained attention, correct and incorrect responses to auditory targets and distractors were used, respectively. By taking into account both hits and false alarms, this score controls for potential response bias, such as a child pressing the space bar in response to each stimulus. Correct responses to the target with a reaction time below 100 milliseconds were excluded (<1% of all trials). The  $d'$  statistic is calculated as follows:  $d' = z(\text{hits}) - z(\text{false alarms})$ . The higher the statistic, the better the child's response sensitivity. Macmillan and Creelman (2005; p. 8) indicate that proportions correct between 0.6 and 0.9 roughly correspond to  $d'$  values between 0.5 and 2.5.

## Data-Analysis

All statistical analyses were done with SPSS 22 (IBM Corp., 2013). Exploration of the data indicated that the dependent variables were normally distributed. NVIQ and SES were added as covariates in all analyses to ensure that these background variables could not influence the results. We first investigated the effects of LI and bilingualism on children's language skills over time, and on their visual and auditory sustained attention measured at wave 1. A  $3 \times 2 \times 2$  mixed-design analysis of covariance (ANCOVA) was conducted for vocabulary and morphology scores separately. Time (Wave 1, 2, and 3) was included as within-subjects factor, and Language Group (monolingual, bilingual) and Impairment Status (TD, LI) as between-subjects factors. *Post hoc* tests were conducted in case significant interactions between the factors in the analyses were observed. For sustained attention, a multivariate ANCOVA included Impairment Status (TD, LI) and Language Group (monolingual, bilingual) as fixed factors and auditory and visual sustained attention as dependent variables. Given the difference in modality, we were hesitant to view the two dependent variables as part of one and the same construct and we thus opted for a MANCOVA instead of a mixed-design ANCOVA (both analyses, however, showed the same patterns).

Subsequently, mediation analyses in the monolingual and bilingual group separately were performed with the PROCESS application for SPSS of Hayes (2013), aiming to find relationships between Impairment Status (the independent variable X), sustained attention (the mediator M), and children's language skills (the dependent variable Y). One important prerequisite of this model is that a cause must precede an effect in time. That is, a change in X must have time to affect a change in M, which, again, must have time to affect a change in Y. To meet the requirement of temporal precedence, we used children's language outcomes



at wave 2 and 3 as dependent variables, and sustained attention measured at wave 1 as mediator. The group distinction (TD-LI), which was the independent variable, was based on assessments prior to wave 1. A visual representation of the mediation model is depicted in **Figure 1**. Separate mediation analyses were done for each language domain at wave 2 and 3 to assess the stability of the effect, and for auditory and visual sustained attention, due to a high correlation between the two ( $r = 0.67$ ,  $p < 0.001$ ). To control for possible effects of language background, all analyses described above were also conducted for a subsample of the participants, excluding bilingual children with LI who had a different first language than the bilingual TD children. Analyses yielded similar results and are therefore not reported.

## RESULTS

### Language Development Vocabulary

**Table 2** presents the means and SDs of children's performance on the PPVT-III-NL, measuring receptive vocabulary. Results revealed a significant main effect of Time [ $F(2,238) = 284.1$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.71$ ], indicating that the vocabulary size of children increased over time, with significant differences across all three waves (all  $p < 0.001$ ). Furthermore, significant main effects of Impairment Status [ $F(1,119) = 33.3$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.22$ ] and Language Group [ $F(1,119) = 26.2$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.18$ ] were found. Children with LI and bilingual children had lower vocabulary scores than TD and monolingual children, respectively. A significant interaction effect of Time  $\times$  Language Group also emerged [ $F(2,238) = 3.1$ ,  $p = 0.047$ ,  $\eta_p^2 = 0.03$ ] and will be discussed below. Other interactions were not significant. Non-verbal IQ was a significant covariate [ $F(1,119) = 18.0$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.13$ ], while SES was not.

*Post hoc* analyses were performed to unpack the interaction between Time  $\times$  Language Group, which showed that the vocabulary size of both monolingual and bilingual children increased over time (all  $p < 0.001$ ). Moreover, univariate ANCOVA's testing group performance on the PPVT-III-NL at wave 1, 2, and 3 separately showed a significant effect of Language Group at each wave. The magnitude of the effect decreased, being large at Wave 1 and medium at Wave 2 and 3 (Wave 1:  $p < 0.001$ ,  $\eta_p^2 = 0.18$ ; Wave 2:  $p < 0.001$ ,  $\eta_p^2 = 0.10$ ; Wave 3:  $p = 0.001$ ,  $\eta_p^2 = 0.09$ ). Thus, the difference in vocabulary size between the monolingual and bilingual children became smaller over time, but the gap was not fully closed.

**TABLE 2 |** Dutch receptive vocabulary skills of the four groups of children (raw scores PPVT).

|      | <i>N</i> <sup>a</sup> | PPVT (raw)<br>Wave 1 |        | PPVT (raw)<br>Wave 2 |        | PPVT (raw)<br>Wave 3 |        |
|------|-----------------------|----------------------|--------|----------------------|--------|----------------------|--------|
|      |                       | Mean (SD)            | Range  | Mean (SD)            | Range  | Mean (SD)            | Range  |
| MOTD | 32                    | 86.6 (10.4)          | 57–103 | 98.5 (9.6)           | 79–122 | 103.8 (8.9)          | 82–124 |
| MOLI | 32                    | 76.5 (9.2)           | 54–93  | 86.6 (11.9)          | 62–110 | 95.9 (10.1)          | 73–114 |
| BITD | 32                    | 75.8 (10.5)          | 57–97  | 87.8 (11.5)          | 64–104 | 96.2 (12.4)          | 70–125 |
| BILI | 32                    | 62.9 (13.3)          | 33–86  | 77.8 (15.5)          | 53–106 | 86.3 (14.0)          | 58–114 |

MOTD, monolingual typically developing; MOLI, monolingual language impaired; BITD, bilingual typically developing; BILI, bilingual language impaired; PPVT, Peabody Picture Vocabulary Task.

<sup>a</sup>For one child in the MOTD group and one child in the BITD group, raw PPVT scores at wave 1 were not available due to incorrect assessment procedures. Moreover, raw PPVT scores at wave 2 were not available for one child in the BITD group.

### Morphology

**Table 3** presents the means and SDs of children's performance on the TAK Word Formation task, measuring grammatical morphology. Results revealed a significant main effect of Time [ $F(2,242) = 167.6$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.58$ ], indicating that children's performance on the word formation task improved over time, with significant differences across all three waves (all  $p < 0.001$ ). In addition, a significant main effect of Impairment Status [ $F(1,121) = 65.8$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.35$ ] and a significant main effect of Language Group [ $F(1,121) = 16.4$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.12$ ] emerged. Children with LI and bilingual children had weaker morphological skills than TD and monolingual children, respectively. There were no significant interaction effects. Non-verbal IQ was a significant covariate [ $F(1,120) = 4.3$ ,  $p = 0.04$ ,  $\eta_p^2 = 0.04$ ], while SES was not.

### Sustained Attention

**Table 4** presents the performance per group on the CPT, split up for auditory and visual stimuli. A multivariate ANCOVA with CPT Auditory and CPT Visual as dependent variables and Impairment Status and Language Group as independent variables revealed a significant negative effect of Impairment

**TABLE 3 |** Dutch morphology skills of the four groups of children (raw scores TAK Word Formation).

|      | <i>N</i> <sup>a</sup> | TAK Word<br>Wave 1 |       | TAK Word<br>Wave 2 |       | TAK Word<br>Wave 3 |       |
|------|-----------------------|--------------------|-------|--------------------|-------|--------------------|-------|
|      |                       | Mean (SD)          | Range | Mean (SD)          | Range | Mean (SD)          | Range |
| MOTD | 32                    | 15.3 (3.8)         | 7–23  | 18.6 (3.5)         | 11–24 | 20.5 (3.6)         | 11–24 |
| MOLI | 32                    | 10.0 (3.7)         | 2–17  | 12.3 (4.1)         | 4–22  | 15.5 (3.9)         | 8–23  |
| BITD | 32                    | 11.7 (5.1)         | 0–20  | 15.3 (4.3)         | 6–21  | 18.1 (3.9)         | 10–24 |
| BILI | 32                    | 6.5 (4.7)          | 0–15  | 11.0 (3.8)         | 0–19  | 12.7 (5.1)         | 3–24  |

MOTD, monolingual typically developing; MOLI, monolingual language impaired; BITD, bilingual typically developing; BILI, bilingual language impaired; TAK, Taaltoets Alle Kinderen.

<sup>a</sup>For one child in the BILI group, raw TAK scores at wave 1 were not available due to a refusal to cooperate.

**TABLE 4 |** Performance on the sustained attention task (CPT d').

|      | N  | CPT (d') Auditory |            | CPT (d') Visual |            |
|------|----|-------------------|------------|-----------------|------------|
|      |    | Mean (SD)         | Range      | Mean (SD)       | Range      |
| MOTD | 32 | 2.33 (0.9)        | 0.53–4.08  | 2.09 (0.8)      | −0.05–3.67 |
| MOLI | 32 | 1.72 (1.0)        | −0.27–3.84 | 1.31 (0.8)      | 0.14–3.63  |
| BITD | 32 | 2.05 (0.8)        | −0.13–4.08 | 2.02 (0.9)      | 0.22–4.08  |
| BILI | 32 | 1.51 (1.0)        | −0.34–3.41 | 1.36 (1.0)      | −0.78–3.67 |

MOTD, monolingual typically developing; MOLI, monolingual language impaired; BITD, bilingual typically developing; BILI, bilingual language impaired; CPT, Continuous Performance Task.

Status [ $F(2,121) = 10.9, p < 0.001, \eta_p^2 = 0.15$ ], whereas there was no main effect of Language Group nor an interaction effect of Impairment Status  $\times$  Language Group. Non-verbal IQ was a significant covariate [ $F(2,121) = 13.5, p < 0.001, \eta_p^2 = 0.18$ ], while SES was not. Bonferroni-corrected pairwise comparisons showed that children with LI scored more poorly on the auditory [ $F(1,122) = 11.2, p = 0.001, \eta_p^2 = 0.08$ ] as well as the visual [ $F(1,122) = 21.4, p < 0.001, \eta_p^2 = 0.15$ ] component of the CPT in comparison with their TD peers. Paired samples  $t$ -tests in each group separately indicated that both the monolingual TD children [ $t(31) = 2.6, p = 0.01, d = 0.28$ ] and the monolingual children with LI [ $t(31) = 2.3, p = 0.03, d = 0.46$ ] performed significantly better on the auditory stimuli than on the visual stimuli. There were no differences between the two components of the CPT in both bilingual groups. Below, mediation analyses investigating the role of auditory and visual sustained attention in explaining the effect of LI on the children's language outcomes will be conducted separately for the monolingual and bilingual group of children.

### Effect of LI in the Monolingual Group

Table 5 presents the results of the mediation analyses investigating the effects of auditory and visual sustained attention on the relation between Impairment Status and language outcomes in the monolingual group of children. To determine whether the effect of Impairment Status on children's language outcomes is significantly reduced due to sustained attention (i.e., the indirect or mediation effect), bootstrapped tests (5,000 – bias-corrected), and confidence intervals were used, as these are more reliable than  $p$ -values. Meaningful mediation is assumed if zero is not included in the confidence intervals of the indirect effects. The results indicate that auditory sustained attention mediated the effect of LI on both vocabulary at wave 2 and 3, and grammatical morphology at wave 2 and 3. At wave 2, the index of mediation (the standardized indirect effect) was slightly larger for vocabulary ( $b = -0.08, 95\% \text{ CI } [-0.22, -0.01]$ ) than morphology ( $b = -0.05, 95\% \text{ CI } [-0.18, -0.001]$ ), but there was substantial overlap in confidence intervals, indicating that reliable differences cannot be assumed. The index of mediation was the same for both domains at wave 3 (vocabulary:  $b = -0.07, 95\% \text{ CI } [-0.21, -0.002]$ ; morphology:  $b = -0.07, 95\% \text{ CI } [-0.19, -0.01]$ ). Although auditory sustained attention significantly reduced the effect of Impairment Status on children's language outcomes, it only accounted for part of the

relationship. The direct effect of Impairment Status on children's language outcomes remained significant when auditory sustained attention was controlled for. Results furthermore showed that visual sustained attention was not a meaningful mediator, as it did not significantly reduce the relation of X on Y. Correlations between children's language and sustained attention skills and visual representations of the mediation models are provided in the Supplementary Table 1 and Figures 1–4, respectively.

### Effect of LI in the Bilingual Group

Table 6 presents the results of the mediation analyses investigating the effects of auditory and visual sustained attention on the relation between Impairment Status and language outcomes in the bilingual group of children. Bootstrapped tests (5,000 – bias-corrected), and confidence intervals were again used to determine whether sustained attention significantly reduced the effect of Impairment Status on vocabulary and morphology. The results from the analyses in the bilingual group suggest that both auditory and visual sustained attention act as partial mediators of the effect of LI on language abilities in both language domains and at both time points. At wave 2, the index of mediation was larger for vocabulary (auditory:  $b = -0.10, 95\% \text{ CI } [-0.25, -0.02]$ ; visual:  $b = -0.14, 95\% \text{ CI } [-0.29, -0.04]$ ) than morphology (auditory:  $b = -0.08, 95\% \text{ CI } [-0.20, -0.02]$ ; visual:  $b = -0.08, 95\% \text{ CI } [-0.22, -0.004]$ ), but there was substantial overlap in confidence intervals, indicating that reliable differences cannot be assumed. At wave 3, the reverse pattern was seen in the analyses with visual sustained attention (vocabulary:  $b = -0.09, 95\% \text{ CI } [-0.23, -0.01]$ ; morphology:  $b = -0.11, 95\% \text{ CI } [-0.25, -0.02]$ ). In the analyses with auditory sustained attention, the index was the same for both domains at wave 3 (vocabulary:  $b = -0.08, 95\% \text{ CI } [-0.20, -0.01]$ ; morphology:  $b = -0.08, 95\% \text{ CI } [-0.22, -0.01]$ ). Correlations between children's language and sustained attention skills and visual representations of the mediation models are provided in the Supplementary Table 2 and Figures 5–8, respectively.

## DISCUSSION

The present study aimed to investigate the effects of an inborn LI and bilingualism on children's language proficiency over time. Moreover, we addressed the question why this child-internal and child-external factor, respectively, produce overlap in children's language profiles (e.g., Paradis, 2005). For the latter, we hypothesized that the language difficulties of children with LI stem from auditory sustained attention deficits, leading to incomplete processing of incoming language. As Leonard (2014) mentioned, "cases of incomplete processing are assumed to be the functional equivalent of reductions in input frequency" (p. 289), which draws a parallel between the origins of the language difficulties of children with LI and bilingual children, whose language skills are influenced by the distributed nature of their language input (Hoff et al., 2012). Two core language domains, i.e., vocabulary and morphology, were chosen as our outcome variables, as these are known to be affected by LI (Krok and Leonard, 2015; Rice and Hoffman, 2015) as well as by reduced

**TABLE 5 |** Mediation effects of auditory and visual sustained attention on the relation between Impairment Status and language outcomes in the monolingual group of children.

| Language outcome | Wave | Effect          | Auditory     |              |              | Visual   |        |       |
|------------------|------|-----------------|--------------|--------------|--------------|----------|--------|-------|
|                  |      |                 | <i>b</i>     | 95% CI       |              | <i>b</i> | 95% CI |       |
|                  |      |                 |              | Lower        | Upper        |          | Lower  | Upper |
| Vocabulary       | 2    | Total effect    | -10.75       | -16.27       | -5.23        | -10.75   | -16.27 | -5.23 |
|                  |      | Direct effect   | -8.79        | -14.29       | -3.30        | -9.45    | -15.53 | -3.36 |
|                  |      | Indirect effect | <b>-1.96</b> | <b>-5.78</b> | <b>-0.14</b> | -1.31    | -5.23  | 1.16  |
|                  | 3    | Total effect    | -6.31        | -11.03       | -1.59        | -6.31    | -11.03 | -1.59 |
|                  |      | Direct effect   | -4.94        | -9.73        | -0.16        | -4.05    | -9.11  | 1.01  |
|                  |      | Indirect effect | <b>-1.36</b> | <b>-4.67</b> | <b>-0.04</b> | -2.26    | -6.59  | 0.07  |
| Morphology       | 2    | Total effect    | -5.83        | -7.82        | -3.85        | -5.83    | -7.82  | -3.85 |
|                  |      | Direct effect   | -5.32        | -7.35        | -3.29        | -5.30    | -7.48  | -3.12 |
|                  |      | Indirect effect | <b>-0.51</b> | <b>-1.74</b> | <b>-0.01</b> | -0.53    | -1.79  | 0.26  |
|                  | 3    | Total effect    | -4.51        | -6.39        | -2.63        | -4.51    | -6.39  | -2.63 |
|                  |      | Direct effect   | -3.87        | -5.76        | -1.99        | -3.91    | -5.97  | -1.85 |
|                  |      | Indirect effect | <b>-0.64</b> | <b>-1.68</b> | <b>-0.06</b> | -0.60    | -1.99  | 0.18  |

CI, confidence interval; Meaningful mediation effects in boldface.

The total effect is the effect of Impairment Status (*X*) on Language (*Y*), excluding Sustained Attention (*M*).

The direct effect is the effect of Impairment Status (*X*) on Language (*Y*), controlling for Sustained Attention (*M*).

The indirect effect is the effect of Impairment Status (*X*) on Language (*Y*) through Sustained Attention (*M*).

input as a result of bilingualism (Scheele et al., 2010; Blom et al., 2012).

With a four-group design, including monolingual and bilingual children with and without LI, we first examined the effects of LI and bilingualism on children's language development in Dutch. Vocabulary and morphology were assessed longitudinally and the results showed that, on both language domains and at each time point, the TD children

outperformed the children with LI and the monolingual children outperformed the bilingual children. These findings correspond to previous work that identified persistent language delays of both children with LI (Rice, 2012; Rice and Hoffman, 2015) and bilingual children (Cobo-Lewis et al., 2002; Farnia and Geva, 2011; Paradis et al., 2016). However, we also found important differences in the way in which LI and (reduced input due to) bilingualism influenced a child's language development. Effects of

**TABLE 6 |** Mediation effects of auditory and visual sustained attention on the relation between Impairment Status and language outcomes in the bilingual group of children.

| Language outcome | Wave | Effect          | Auditory     |              |              | Visual       |              |              |
|------------------|------|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                  |      |                 | <i>b</i>     | 95% CI       |              | <i>b</i>     | 95% CI       |              |
|                  |      |                 |              | Lower        | Upper        |              | Lower        | Upper        |
| Vocabulary       | 2    | Total effect    | -9.82        | -16.22       | -3.42        | -9.82        | -16.22       | -3.42        |
|                  |      | Direct effect   | -7.08        | -13.43       | -0.73        | -6.15        | -12.57       | 0.26         |
|                  |      | Indirect effect | <b>-2.74</b> | <b>-6.67</b> | <b>-0.62</b> | <b>-3.67</b> | <b>-8.15</b> | <b>-0.94</b> |
|                  | 3    | Total effect    | -10.02       | -16.36       | -3.68        | -10.02       | -16.36       | -3.68        |
|                  |      | Direct effect   | -7.95        | -14.37       | -1.53        | -7.71        | -14.34       | -1.08        |
|                  |      | Indirect effect | <b>-2.07</b> | <b>-5.55</b> | <b>-0.29</b> | <b>-2.31</b> | <b>-6.25</b> | <b>-0.26</b> |
| Morphology       | 2    | Total effect    | -4.44        | -6.48        | -2.40        | -4.44        | -6.48        | -2.40        |
|                  |      | Direct effect   | -3.76        | -5.82        | -1.70        | -3.67        | -5.80        | -1.55        |
|                  |      | Indirect effect | <b>-0.68</b> | <b>-1.83</b> | <b>-0.13</b> | <b>-0.77</b> | <b>-2.17</b> | <b>-0.03</b> |
|                  | 3    | Total effect    | -5.57        | -7.77        | -3.38        | -5.57        | -7.77        | -3.38        |
|                  |      | Direct effect   | -4.77        | -6.97        | -2.57        | -4.48        | -6.71        | -2.25        |
|                  |      | Indirect effect | <b>-0.80</b> | <b>-2.26</b> | <b>-0.14</b> | <b>-1.09</b> | <b>-2.75</b> | <b>-0.23</b> |

CI, confidence interval; Meaningful mediation effects in boldface.

The total effect is the effect of Impairment Status (*X*) on Language (*Y*), excluding Sustained Attention (*M*).

The direct effect is the effect of Impairment Status (*X*) on Language (*Y*), controlling for Sustained Attention (*M*).

The indirect effect is the effect of Impairment Status (*X*) on Language (*Y*) through Sustained Attention (*M*).

LI on vocabulary and morphology were large and remained stable over time, as expected (Rice, 2012). The effect of bilingualism on morphology also remained stable over time, likely due to a number of irregular items in our morphology task which have a low type frequency and are typically acquired at a late age (see, Boerma et al., 2017), but this effect was considerably smaller in magnitude than the effect of LI. Moreover, the difference in vocabulary size between the monolingual and bilingual children diminished over time, like in Farnia and Geva (2011). Despite persistent language delays in both groups, the most extensive overlap between the language profiles of the children with LI and bilingual children was thus evident on vocabulary in early (pre)school years. Future longitudinal research covering a longer period of time is needed to examine whether the overlap further reduces in later developmental stages.

To understand the source of this overlap, we furthermore investigated the effects of LI and bilingualism on children's auditory and visual sustained attention skills, and explored the role of sustained attention in explaining the effects of LI on children's language outcomes. In accordance with the meta-analysis of Ebert and Kohnert (2011), we found that the children with LI had a weaker ability to maintain their attention to the auditory and visual stimuli of the CPT than the TD children. Contrary to our predictions, the children with LI did not have more extensive problems with the auditory than the visual stimuli. Instead, the monolingual children with LI, like their monolingual TD peers, showed the reverse pattern, with a better performance on the auditory component of the CPT. This finding may be related to the integrated set-up of our task, in which auditory and visual stimuli were interspersedly presented during a prolonged period of time. To accurately respond to the visual targets and distractors, children were required to stay focused on the computer screen, whereas a quick look in another direction did not necessarily affect responses to auditory stimuli. Interestingly, this task effect did not influence the sustained attention performance of the bilingual children, both TD and LI, whose response sensitivity on the two modalities did not differ. It would be worthwhile to examine whether the use of a different sustained attention measure, with separate blocks of only visual or only auditory stimuli, would show the same results. We will come back to the discrepancy between the monolingual and bilingual children when discussing the outcomes of the mediation analyses.

While the results showed that LI was associated with weak sustained attention, no effect of bilingualism was found. Monolingual and bilingual participants scored equally well on the auditory and visual components of the CPT. Previous work reported a bilingual advantage on different attention measures (e.g., Bialystok, 1999; Engel de Abreu et al., 2012), but, to our knowledge, the current study is the first to specifically investigate sustained attention in bilingual children. Although Krizman et al. (2012) found better performance of bilingual adults in comparison with monolingual adults on a task targeting sustained attention, other adult studies failed to find this specific advantage (Bialystok et al., 2008; Bak et al., 2014). There are several factors that have been shown to moderate the effect of

bilingualism on attention (and other aspects of cognition), which may explain the mixed findings in the literature and the absent positive effect of bilingualism in the current study. For example, a number of studies have shown that cognitive advantages are limited to bilinguals who are proficient in both languages (Carlson and Meltzoff, 2008; Poarch and van Hell, 2012; Weber et al., 2016) or emerge as an effect of growing bilingual proficiency (Blom et al., 2014; Crivello et al., 2016). It may thus be that the language proficiency of the bilingual children in our sample was not sufficiently strong for cognitive advantages to develop. In addition, it is also conceivable that bilinguals benefit from their bilingual language experience on certain cognitive measures, but not on others, as Bialystok et al. (2008) argue. Although common measures for sustained attention (including the measure used in the present study) require a degree of response inhibition, they involve simple stimuli and a rule dictating when to respond or refrain from responding. In contrast, measures such as the Simon or Stroop task, which also tap into attentional processing and on which a bilingual advantage has traditionally been found, use complex stimuli with multiple features that include a salient conflict (direction vs. position or word vs. color). Such conflict-monitoring is trained by interactions in bilingual contexts, explaining why a bilingual benefit may be limited to tasks that require substantial conflict resolution (for an elaborate discussion, see Bialystok et al., 2008). Nevertheless, even on those measures that require substantial conflict resolution, bilingual advantages are not always found (e.g., Antón et al., 2014; Duñabeitia et al., 2014), indicating that it is yet unclear under which specific conditions a bilingual benefit emerges.

To explore relations between the poor language abilities and the poor sustained attention skills of children with LI, we performed mediation analyses. Results showed that auditory sustained attention mediated the effect of LI on children's language outcomes. This effect was stable, emerging on vocabulary and morphology, at wave 2 and 3, in the monolingual and bilingual group. These findings are in line with previous research that indicated positive associations between language and sustained attention in children with LI (Montgomery, 2008; Montgomery et al., 2009; Duinmeijer et al., 2012; Ebert et al., 2012, 2014; Blom and Boerma, 2016; Jongman et al., 2016). Although we hypothesized that sustained attention effects would be more pronounced on vocabulary than morphology, as a result of their susceptibility to input effects (Chondrogianni and Marinis, 2011), reliable differences between the two language domains were not found. As was mentioned before, this may be due to the complex irregular structures included in our morphology task (see, Gathercole, 2010; Paradis, 2010a). The inclusion of only regular items could possibly lead to different results and is an interesting venue for future research. Contrary to auditory sustained attention, visual sustained attention did not act as a meaningful mediator of the effect of LI on monolingual children's language skills. This contrast between the auditory and visual modality seems to confirm our hypothesis that the language difficulties of the monolingual children with LI reflect, at least in part, a domain-specific weakened ability to maintain attention to auditory information, leading to

incomplete processing of incoming language input. Thus, while reductions in input frequency cause language delays in bilingual children, the functional equivalent may impair the language proficiency of children with LI, resulting in partially overlapping language profiles.

In contrast to the monolingual children and contrary to our expectations, visual sustained attention did mediate the effect of LI on the vocabulary and morphology scores of bilingual children. Moreover, as mentioned before, there was also a discrepancy between the monolinguals and the bilinguals in terms of relative performance on the visual and auditory components of the CPT. While the two monolingual groups of children scored better on the auditory than the visual stimuli, the two bilingual groups performed equally well on both modalities. These discrepancies in our findings between the monolinguals and bilinguals may be related to research which showed that bilingual children attend more to visual speech cues in the environment in comparison with monolingual children, for whom these cues are redundant (Pons et al., 2015). In support of the complex task of dual language acquisition, bilinguals may exploit such visual information during social interactions more than monolinguals, enhancing the importance of visual sustained attention for successful language learning in bilingual contexts. If a child is less able to make use of these visual cues, due to poor visual sustained attention, this will hinder their acquisition of language, which is what the results from the present study suggest. Another possibility is that bilingual children rely more on orthographic learning than monolingual children to boost their second language skills. Several studies have shown that vocabulary learning in different populations, including bilinguals (Vadasy and Sanders, 2016) and children with LI (Ricketts et al., 2015), benefits from the presence of orthography. It may be that these orthographic facilitation effects are particularly strong in the context of dual language learning, explaining why visual sustained attention mediated the effect of LI on language in the bilingual group of children. Future research is necessary to investigate this hypothesis.

An alternative explanation for our findings could be that relations between children's poor language abilities and poor sustained attention skills emerged as a result of a task effect. It may be that children need sustained attention to successfully complete the vocabulary and morphology task that we used to assess language competence. While this alternative interpretation cannot be ruled out, it does not accurately explain the discrepancy in our results between the auditory and the visual domain in the monolingual group of children. During both the vocabulary and the morphology task, children were required to maintain their attention to pictures as well as verbally presented words or sentences. If our findings were a mere reflection of task effects, both visual and auditory sustained attention would be expected to play a role. To investigate if attention influences children's language performance in a task or also their language learning process, follow-up research could consider using measures from spontaneous speech data or using an experimental paradigm in which attention load is manipulated.

Although the findings from this study point to the importance of attention resources for the language proficiency of children with LI, they also indicate that sustained attention deficits only accounted for part of the effect of LI on children's language skills. This is not surprising, as LI is a complex multifaceted disorder with no single underlying cause (Bishop, 2006). Future research is recommended to investigate multiple cognitive risk factors of LI, for example including both sustained attention and working memory, considering their individual contributions to the language deficit as well as how they interact. Moreover, future work needs to study the bidirectional relationships between language and cognition to further understand the behavioral profile of children with LI. The current study explored the effect of cognition on language, but reverse influences of language proficiency on cognition are also likely (e.g., Fuhs and Day, 2011; Kuhn et al., 2016) and could explain the co-occurrence of linguistic and non-linguistic weaknesses of children with LI (but see, Gooch et al., 2016). Finally, this study was limited by the heterogeneous sample of bilingual children, restricting the possibility to draw conclusions about specific groups. The bilingual children in our sample all learned Dutch as a second language, but varied considerably in degrees of exposure to Dutch and first language background. Such factors influence the severity and persistence of a bilingual child's language delay (e.g., Paradis, 2010b; Blom et al., 2012), and are important to take into account in future work.

## CONCLUSION

The current study provided insight into the persistence and origins of the partially overlapping language profiles of bilingual children and children with LI. Our results showed that the language abilities of bilingual children and children with LI were persistently weaker than the language skills of monolingual and TD children, respectively. The overlap between the language profiles of bilingual children and children with LI was particularly large for vocabulary in early (pre)school years and diminished over time. Furthermore, our findings indicate that the overlap may be explained by the weakened ability of children with LI to maintain attention to the stream of linguistic information, interfering with how well incoming language is processed. While reductions in input frequency cause language delays in bilingual children, the functional equivalent, i.e., incomplete processing of input, may impair the language proficiency of children with LI. Next to auditory sustained attention, visual sustained attention also partly accounted for the language difficulties of bilingual children with LI, in contrast to their monolingual peers. These outcomes prompt further research on relations between LI, language skills and cognition in both monolingual and bilingual learning settings.

## ETHICS STATEMENT

This research was screened by the Standing Ethical Assessment Committee of the Faculty of Social and Behavioral Sciences at Utrecht University. Criteria were met and further verification

was not deemed necessary. Parents of participants gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

All authors were involved in the conception and design of the study. TB wrote the manuscript and conducted the statistical analyses. PL, FW, and EB revised the draft for critical content.

## FUNDING

This work is part of the research program ‘Cognitive development in the context of emerging bilingualism: Cultural minority children in the Netherlands’ which is financed by a VIDI-grant

awarded to EB by the Netherlands Organization for Scientific Research (NWO; grant number 016.124.369).

## ACKNOWLEDGMENT

We thank the children, parents, and schools that participated in the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01241/full#supplementary-material>

## REFERENCES

- Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., et al. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Front. Psychol.* 5:398. doi: 10.3389/fpsyg.2014.00398
- Appel, R., and Vermeer, A. (1998). Speeding up second language vocabulary acquisition of minority children. *Lang. Educ.* 12, 159–173. doi: 10.1080/09500789808666746
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423.
- Bak, T. H., Vega-Mendoza, M., and Sorace, A. (2014). Never too late? An advantage on tests of auditory attention extends to late bilinguals. *Front. Psychol.* 5:485. doi: 10.3389/fpsyg.2014.00485
- Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., and Lancee, W. (1996). Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. *J. Am. Acad. Child Adolesc. Psychiatry* 35, 804–814. doi: 10.1097/00004583-199606000-00021
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Dev.* 70, 636–644. doi: 10.1111/1467-8624.00046
- Bialystok, E., Craik, F., and Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *J. Exp. Psychol.* 34, 859–873. doi: 10.1037/0278-7393.34.4.859
- Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism* 13, 525–531. doi: 10.1017/S1366728909990423
- Bishop, D. V. (2006). What causes specific language impairment in children? *Curr. Dir. Psychol. Sci.* 15, 217–221. doi: 10.1111/j.1467-8721.2006.00439.x
- Blom, E., and Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Res. Dev. Disabil.* 55, 301–311. doi: 10.1016/j.ridd.2016.05.001
- Blom, E., and Boerma, T. (2017). Effects of language impairment and bilingualism across domains: vocabulary, morphology and verbal memory. *Linguist. Approaches Biling.* 7:3/4, 277–300. doi: 10.1075/lab.15018.blo
- Blom, E., de Jong, J., Orgassa, A., Baker, A., and Weerman, F. (2013). Verb inflection in monolingual Dutch and sequential bilingual Turkish-Dutch children with and without SLI. *Int. J. Lang. Commun. Disord.* 48, 382–393. doi: 10.1111/1460-6984.12013
- Blom, E., Küntay, A. C., Messer, M., Verhagen, J., and Leseman, P. (2014). The benefits of being bilingual: working memory in bilingual Turkish-Dutch children. *J. Exp. Child Psychol.* 128, 105–119. doi: 10.1016/j.jecp.2014.06.007
- Blom, E., Paradis, J., and Duncan, T. S. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular-s in child L2 english. *Lang. Learn.* 62, 965–994. doi: 10.1111/j.1467-9922.2012.00715.x
- Boerma, T., Wijnen, F., Leseman, P., and Blom, E. (2017). Grammatical morphology in monolingual and bilingual children with and without language impairment: the case of Dutch plurals and past participles. *J. Speech Lang. Hear. Res.* 60, 2064–2080. doi: 10.1044/2017\_jslhr-16-0351
- Carlson, S. M., and Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci.* 11, 282–298. doi: 10.1111/j.1467-7687.2008.00675.x
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, morphology and complex syntax in successive bilingual children. *Linguist. Approaches Biling.* 1, 223–248. doi: 10.1075/lab.1.3.05cho
- Cobo-Lewis, A., Pearson, B., Eilers, R., and Umbel, V. (2002). “Effects of bilingualism and bilingual education on oral and written English skills: a multifactor study of standardized test outcomes,” in *Language and Literacy in Bilingual Children*, eds D. K. Oller and R. Eilers (Clevedon: Multilingual Matters), 64–97.
- Conti-Ramsden, G. (2010). Summing up problems in bilingual specific language impairment: why multiple influences may not be additive. *Appl. Psycholinguist.* 31, 270–273. doi: 10.1017/S0142716409990403
- Conti-Ramsden, G., St Clair, M. C., Pickles, A., and Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: from childhood to adolescence. *J. Speech Lang. Hear. Res.* 55, 1716–1735. doi: 10.1044/1092-4388(2012/10-0182)
- Conti-Ramsden, G., Ullman, M. T., and Lum, J. A. (2015). The relation between receptive grammar and procedural, declarative, and working memory in specific language impairment. *Front. Psychol.* 6:1090. doi: 10.3389/fpsyg.2015.01090
- Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford Psychology Series, No. 26. New York, NY: Oxford University Press.
- Crivello, C., Kuzyk, O., Rodrigues, M., Friend, M., Zesiger, P., and Poulin-Dubois, D. (2016). The effects of bilingual growth on toddlers’ executive function. *J. Exp. Child Psychol.* 141, 121–132. doi: 10.1016/j.jecp.2015.08.004
- Dodwell, K., and Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *Int. J. Lang. Commun. Disord.* 43, 201–218. doi: 10.1080/13682820701366147
- Duinmeijer, I., de Jong, J., and Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *Int. J. Lang. Commun. Dis.* 47, 542–555. doi: 10.1111/j.1460-6984.2012.00164.x
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., et al. (2014). The inhibitory advantage in bilingual children revisited. *Exp. Psychol.* 61, 234–251. doi: 10.1027/1618-3169/a000243
- Ebert, K. D., and Kohnert, K. (2011). Sustained attention in children with primary language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 54, 1372–1384. doi: 10.1044/1092-4388(2011/10-0231)
- Ebert, K. D., Kohnert, K., Pham, G., Rentmeester-Disher, J. R., and Payesteh, B. (2014). Three treatments for bilingual children with primary language impairment: examining cross-linguistic and cross-domain effects. *J. Speech Lang. Hear. Res.* 57, 172–186. doi: 10.1044/1092-4388(2013/12-0388)
- Ebert, K. D., Rentmeester-Disher, J., and Kohnert, K. (2012). Nonlinguistic cognitive treatment for bilingual children with primary language impairment. *Clin. Ling. Phonet.* 26, 485–501. doi: 10.3109/02699206.2012.660226

- Engel de Abreu, P. M. E., Cruz-Santos, A., Tourinho, C. J., Martin, R., and Bialystok, E. (2012). Bilingualism enriches the poor enhanced cognitive control in low-income minority children. *Psychol. Sci.* 23, 1364–1371. doi: 10.1177/0956797612443836
- Farnia, F., and Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Appl. Psycholinguist.* 32, 711–738. doi: 10.1017/S0142716411000038
- Finneran, D. A., Francis, A. L., and Leonard, L. B. (2009). Sustained attention in children with specific language impairment (SLI). *J. Speech Lang. Hear. Res.* 52, 915–929. doi: 10.1044/1092-4388(2009/07-0053)
- Fuhs, M. W., and Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Dev. Psychol.* 47, 404–416. doi: 10.1037/a0021065
- Gathercole, S. E. (2006). Nonword repetition and word learning: the nature of the relationship. *Appl. Psycholinguist.* 27, 513–543. doi: 10.1017/s0142716406060383
- Gathercole, V. C. M. (2010). Interface or face to face? The profiles and contours of bilinguals and specific language impairment. *Appl. Psycholinguist.* 31, 282–293. doi: 10.1017/S0142716409990439
- Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., and Cowan, N. (2000). The development of auditory attention in children. *Front. Biosci.* 5:d108–d120. doi: 10.2741/a500
- Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., and Hulme, C. (2016). The development of executive function and language skills in the early school years. *J. Child Psychol. Psychiatry* 57, 180–187. doi: 10.1111/jcpp.12458
- Gray, S. (2003). Word-learning by preschoolers with specific language impairment: what predicts success? *J. Speech Lang. Hear. Res.* 46, 56–67. doi: 10.1044/1092-4388(2003/005)
- Grüter, T. (2005). Comprehension and production of French object clitics by child second language learners and children with specific language impairment. *Appl. Psycholinguist.* 26, 363–391. doi: 10.1017/S0142716405050216
- Grüter, T., and Paradis, J. (eds) (2014). *Input and Experience in Bilingual Development*, Vol. 13. Amsterdam: John Benjamins Publishing Company.
- Gutiérrez-Clellen, V., Simon-Cerejido, G., and Wagner, C. (2008). Bilingual children with language impairment: a comparison with monolinguals and second language learners. *Appl. Psycholinguist.* 29, 3–20. doi: 10.1017/S0142716408080016
- Hari, R., and Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends Cogn. Sci.* 5, 525–532. doi: 10.1016/S1364-6613(00)01801-5
- Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experiences of Young American Children*. Baltimore, MD: Brookes Publishing.
- Hayes, A. F. (2013). *An Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.
- Henry, L. A., and Botting, N. (2016). Working memory and developmental language impairments. *Child Lang. Teach. Ther.* 33, 19–32. doi: 10.1177/0265659016655378
- Henry, L. A., Messer, D. J., and Nash, G. (2012). Executive functioning in children with specific language impairment. *J. Child Psychol. Psychiatry* 53, 37–45. doi: 10.1111/j.1469-7610.2011.02430.x
- Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., and Levine, S. (2002). Language input and child syntax. *Cognit. Psychol.* 45, 337–374. doi: 10.1016/S0010-0285(02)00500-5
- Jia, G., and Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: age-related differences. *J. Speech Lang. Hear. Res.* 50, 1280–1299. doi: 10.1044/1092-4388(2007/090)
- Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., et al. (1999). Fourteen-year follow-up of children with and without speech/language impairments speech/language stability and outcomes. *J. Speech Lang. Hear. Res.* 42, 744–760. doi: 10.1044/jslhr.4203.744
- Jongman, S. R., Roelofs, A., Scheper, A. R., and Meyer, A. S. (2016). Picture naming in typically developing and language-impaired children: the role of sustained attention. *Int. J. Lang. Commun. Disord.* 52, 323–333. doi: 10.1111/1460-6984.12275
- Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *J. Speech Lang. Hear. Res.* 37, 418–421. doi: 10.1044/jslhr.3702.418
- Kan, P. F., and Windsor, J. (2010). Word learning in children with primary language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 53, 739–756. doi: 10.1044/1092-4388(2009/08-0248)
- Kapa, L. L., and Colombo, J. (2014). Executive function predicts artificial language learning. *J. Mem. Lang.* 76, 237–252. doi: 10.1016/j.jml.2014.07.004
- Kapa, L. L., and Plante, E. (2015). Executive function in SLI: recent advances and future directions. *Curr. Dev. Disord. Rep.* 2, 245–252. doi: 10.1007/s40474-015-0050-x
- Kort, W., Schittekatte, M., and Compaan, E. L. (2008). *CELF-4-NL: Clinical Evaluation of Language Fundamentals*. Amsterdam: Pearson Assessment and Information B.V.
- Krizman, J., Marian, V., Shook, A., Skoe, E., and Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proc. Natl. Acad. Sci. U.S.A.* 109, 7877–7881. doi: 10.1073/pnas.1201575109
- Krok, W. C., and Leonard, L. B. (2015). Past tense production in children with and without specific language impairment across Germanic languages: a meta-analysis. *J. Speech Lang. Hear. Res.* 58, 1326–1340. doi: 10.1044/2015\_JSLHR-L-14-0348
- Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., Blair, C. B., and Family Life Project Key Investigators (2016). The contribution of children's time-specific and longitudinal expressive language skills on developmental trajectories of executive function. *J. Exp. Child Psychol.* 148, 20–34. doi: 10.1016/j.jecp.2016.03.008
- Law, J., Tomblin, J. B., and Zhang, X. (2008). Characterizing the growth trajectories of language-impaired children between 7 and 11 years of age. *J. Speech Lang. Hear. Res.* 51, 739–749. doi: 10.1044/1092-4388(2008/052)
- Leonard, L. B. (2014). *Children with Specific Language Impairment*. Cambridge: MIT Press.
- Leonard, L. B., Davis, J., and Deevy, P. (2007a). Phonotactic probability and past tense use by children with specific language impairment and their typically developing peers. *Clin. Ling. Phonet.* 21, 747–758.
- Leonard, L. B., Eyer, J. A., Bedore, L. M., and Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *J. Speech Lang. Hear. Res.* 40, 741–753. doi: 10.1044/jslhr.4004.741
- Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., and Kail, R. V. (2007b). Speed of processing, working memory, and language impairment in children. *J. Speech Lang. Hear. Res.* 50, 408–428.
- Leroy, S., Parisse, C., and Maillart, C. (2013). The influence of the frequency of functional markers on repetitive imitation of syntactic constructions in children with specific language impairment, from their own language productions. *Clin. Ling. Phonet.* 27, 508–520. doi: 10.3109/02699206.2013.787546
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide*. New York, NY: Cambridge University Press.
- Marton, K. (2008). Visuo-spatial processing and executive functions in children with specific language impairment. *Int. J. Lang. Commun. Disord.* 43, 181–200. doi: 10.1080/16066350701340719
- Miller, C. A., Kail, R., Leonard, L. B., and Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *J. Speech Lang. Hear. Res.* 44, 416–433. doi: 10.1044/1092-4388(2001/034)
- Montgomery, J. W. (2008). Role of auditory attention in the real-time processing of simple grammar by children with specific language impairment: a preliminary investigation. *Int. J. Lang. Commun. Disord.* 43, 499–527. doi: 10.1080/13682820701736638
- Montgomery, J. W., Evans, J. L., and Gillam, R. B. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: a preliminary study. *Appl. Psycholinguist.* 30, 123–151. doi: 10.1017/S0142716408090061
- Montgomery, J. W., Magimairaj, B. M., and Finney, M. C. (2010). Working memory and specific language impairment: an update on the relation and perspectives on assessment and treatment. *Am. J. Speech Lang. Pathol.* 19, 78–94. doi: 10.1044/1058-0360(2009/09-0028)
- Noterdaeme, M., Amorosa, H., Mildenberger, K., Sitter, S., and Minow, F. (2001). Evaluation of attention problems in children with autism and children

- with a specific language disorder. *Eur. Child Adolesc. Psychiatry* 10, 58–66. doi: 10.1007/s007870170048
- Paap, K. R., Johnson, H. A., and Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex* 69, 265–278. doi: 10.1016/j.cortex.2015.04.014
- Paradis, J. (2005). Grammatical morphology in children learning English as a second language implications of similarities with specific language impairment. *Lang. Speech Hear. Serv. Sch.* 36, 172–187. doi: 10.1044/0161-1461(2005/019)
- Paradis, J. (2007). Bilingual children with specific language impairment: theoretical and applied issues. *Appl. Psycholinguist.* 28, 551–564. doi: 10.1017/S0142716407070300
- Paradis, J. (2010a). Bilingual children's acquisition of English verb morphology: effects of language exposure, structure complexity, and task type. *Lang. Learn.* 60, 651–680. doi: 10.1111/j.1467-9922.2010.00567.x
- Paradis, J. (2010b). The interface between bilingual development and specific language impairment. *Appl. Psycholinguist.* 31, 3–28. doi: 10.1017/s0142716409990373
- Paradis, J. (2011). Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguist. Approaches Biling.* 1, 213–237. doi: 10.1075/lab.1.3.01par
- Paradis, J., and Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Dev. Sci.* 20:e12433. doi: 10.1111/desc.12433
- Paradis, J., Jia, R., and Arppe, A. (2017). The acquisition of tense morphology over time by English second language children with specific language impairment: testing the cumulative effects hypothesis. *Appl. Psycholinguist.* 38, 881–908. doi: 10.1017/S0142716416000485
- Paradis, J., Tulp, Y., and Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: individual variation in long-term outcomes. *J. Child Lang.* 43, 553–580. doi: 10.1017/S0305000915000562
- Poarch, G. J., and van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *J. Exp. Child Psychol.* 113, 535–551. doi: 10.1016/j.jecp.2012.06.013
- Pons, F., Bosch, L., and Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychol. Sci.* 26, 490–498. doi: 10.1177/0956797614568320
- Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., et al. (2014). Specific language impairment: a convenient label for whom? *Int. J. Lang. Commun. Disord.* 49, 416–451. doi: 10.1111/1460-6984.12102
- Rice, M., Wexler, K., and Hershberger, S. (1998). Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. *J. Speech Lang. Hear. Res.* 41, 1412–1431. doi: 10.1044/jslhr.41.06.1412
- Rice, M. L. (2004). "Growth models of developmental language disorders," in *Developmental Language Disorders: From Phenotypes to Etiologies*, eds M. L. Rice and S. F. Warren (Mahwah, NJ: Lawrence Erlbaum Associates), 207–240.
- Rice, M. L. (2012). Toward epigenetic and gene regulation models of specific language impairment: looking for links among growth, genes, and impairments. *J. Neurodev. Disord.* 4:27. doi: 10.1186/1866-1955-4-27
- Rice, M. L., and Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: a longitudinal study from 2;6 to 21 years of age. *J. Speech Lang. Hear. Res.* 58, 345–359. doi: 10.1044/2015\_JSLHR-L-14-0150
- Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., and Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *J. Speech Lang. Hear. Res.* 37, 106–122. doi: 10.1044/jslhr.37.01.106
- Riches, N. G., Tomasello, M., and Conti-Ramsden, G. (2005). Verb learning in children with SLI. Frequency and spacing effects. *J. Speech Lang. Hear. Res.* 48, 1397–1411. doi: 10.1044/1092-4388(2005/097)
- Ricketts, J., Dockrell, J. E., Patel, N., Charman, T., and Lindsay, G. (2015). Do children with specific language impairment and autism spectrum disorders benefit from the presence of orthography when learning new spoken words? *J. Exp. Child Psychol.* 134, 43–61. doi: 10.1016/j.jecp.2015.01.015
- Rispens, J., and de Bree, E. (2015). Bilingual children's production of regular and irregular past tense morphology. *Bilingualism* 18, 290–303. doi: 10.1017/S1366728914000108
- Roessingh, H., and Elgie, S. (2009). Early language and literacy development among young English language learners: preliminary insights from a longitudinal study. *TESL Canada J.* 26, 24–45. doi: 10.18806/tesl.v26i2.413
- Rothweiler, M., Chilla, S., and Clahsen, H. (2012). Subject-verb agreement in specific language impairment: a study of monolingual and bilingual German-speaking children. *Bilingualism* 15, 39–57. doi: 10.3109/02699206.2014.886726
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev.* 83, 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x
- Sandford, J. A., and Turner, A. (2004). *IVA + plus: Integrated Visual and Auditory Continuous Performance Test Administration Manual*. Richmond, VA: BrainTrain.
- Scarborough, H., and Dobrich, W. (1990). Development of children with early language delay. *J. Speech Lang. Hear. Res.* 33, 70–83. doi: 10.1044/jslhr.33.01.70
- Scheele, A. F. (2010). *Home Language and Mono-and Bilingual Children's Emergent Academic Language: A Longitudinal Study of Dutch, Moroccan-Dutch, and Turkish-Dutch 3-to 6-Year-Old Children*. Enschede: Ipskamp Drukkers.
- Scheele, A. F., Leseman, P. P., and Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Appl. Psycholinguist.* 31, 117–140. doi: 10.1017/S0142716409990191
- Schlichting, J. E. P. T., and Lutje Spelberg, H. C. (2010a). *Schlichting Test Voor Taalbegrip-I: Manual*. Houten: Bohn, Stafleu, Van Loghum.
- Schlichting, J. E. P. T., and Lutje Spelberg, H. C. (2010b). *Schlichting Test Voor Taalproductie-II: Manual*. Houten: Bohn, Stafleu, Van Loghum.
- Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL. Dutch version*. Amsterdam: Harcourt Assessment B.V.
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools Inc.
- Silvén, M., Voeten, M., Kouvo, A., and Lundén, M. (2014). Speech perception and vocabulary growth: a longitudinal study of Finnish-Russian bilinguals and Finnish monolinguals from infancy to three years. *Int. J. Behav. Dev.* 38, 323–332. doi: 10.1177/0165025414533748
- Spaulding, T. J., Plante, E., and Vance, R. (2008). Sustained selective attention skills of preschool children with specific language impairment: evidence for separate attentional capacities. *J. Speech Lang. Hear. Res.* 51, 16–34. doi: 10.1044/1092-4388(2008/002)
- Stevens, C., and Bavelier, D. (2012). The role of selective attention on academic foundations: a cognitive neuroscience perspective. *Dev. Cogn. Neurosci.* 2, S30–S48. doi: 10.1016/j.dcn.2011.11.001
- Stichting Siméa (2014). *Indicatiecriteria: Auditief en/of Communicatief Beperkte Leerlingen*. Available at: <http://www.simea.nl/dossiers/passend-onderwijs/brochures-po/simea-brochure-indicatiecriteria-juni-2014.pdf>
- Stichting Siméa (2016). *Handreiking Meertaligheid TOS*. Available at: <http://www.simea.nl/dossiers/si166289-simea-handreiking-meertaligheid-tos.pdf>
- Stromswold, K. (2001). The heritability of language: a review and meta analysis of twin, adoption, and linkage studies. *Language* 77, 647–723. doi: 10.1353/lan.2001.0247
- Thomas, E. M., Williams, N., Jones, L. A., Davies, S., and Binks, H. (2014). Acquiring complex structures under minority language conditions: bilingual acquisition of plural morphology in welsh. *Bilingualism* 17, 478–494. doi: 10.1017/S1366728913000497
- Thordardottir, E., and Juliusdottir, A. (2013). Icelandic as a second language: a longitudinal study of language knowledge and processing by school-age children. *Int. J. Biling. Educ. Biling.* 16, 411–435. doi: 10.1080/13670050.2012.693062
- Thordardottir, E., Rothenberg, A., Rivard, M., and Naves, R. (2006). Bilingual assessment: can overall proficiency be estimated from separate measurement of two languages? *J. Multiling. Commun. Disord.* 4, 1–21. doi: 10.1080/14769670500215647
- Tirosh, E., and Cohen, A. (1998). Language deficit with attention-deficit disorder: a prevalent comorbidity. *J. Child Neurol.* 13, 493–497. doi: 10.1177/088307389801301005
- Tuller, L. (2015). "Clinical use of parental questionnaires in multilingual contexts," in *Methods for Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. De Jong, and N. Meir (Bristol: Multilingual Matters), 301–330.

- Ullman, M. T., and Pierpont, E. I. (2005). Specific language impairment is not specific to language: the procedural deficit hypothesis. *Cortex* 41, 399–433. doi: 10.1016/S0010-9452(08)70276-4
- Vadasy, P. F., and Sanders, E. A. (2016). Attention to orthographic and phonological word forms in vocabulary instruction for kindergarten English learners. *Read. Psychol.* 37, 833–866. doi: 10.1080/02702711.2015.1116477
- Vagh, S. B., Pan, B. A., and Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: the utility of combining parent and teacher report. *Child Dev.* 80, 1545–1563. doi: 10.1111/j.1467-8624.2009.01350.x
- Verhoeven, L., Steenge, J., van Weerdenburg, M., and van Balkom, H. (2011). Assessment of second language proficiency in bilingual children with specific language impairment: a clinical perspective. *Res. Dev. Disabil.* 32, 1798–1807. doi: 10.1016/j.ridd.2011.03.010
- Verhoeven, L., and Vermeer, A. (2001). *Taaltoets Alle Kinderen (TAK)*. Arnhem: Cito.
- Vissers, C., Koolen, S., Hermans, D., Scheper, A., and Knoors, H. (2015). Executive functioning in preschoolers with specific language impairment. *Front. Psychol.* 6:1574. doi: 10.3389/fpsyg.2015.01574
- Weber, R. C., Johnson, A., Riccio, C. A., and Liew, J. (2016). Balanced bilingualism and executive functioning in children. *Bilingualism* 19, 425–431. doi: 10.1093/arclin/acq097
- Wechsler, D., and Naglieri, J. A. (2008). *Wechsler Nonverbal Scale of Ability*. Available at: <http://www.pearsonclinical.nl/wnv-nl-wechsler-non-verbal>
- Yoshida, H., Tran, D. N., Benitez, V., and Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Front. Psychol.* 2:210. doi: 10.3389/fpsyg.2011.00210

**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 © 2017 Boerma, Leseman, Wijnen and Blom. 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.



# Methods for Identifying Specific Language Impairment in Bilingual Populations in Germany

Cornelia Hamann\* and Lina Abed Ibrahim

Department of English, University of Oldenburg, Oldenburg, Germany

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Vicky Chondrogianni,  
University of Edinburgh,  
United Kingdom  
Fabrizio Arosio,  
University of Milano-Bicocca, Italy

### \*Correspondence:

Cornelia Hamann  
cornelia.hamann@uni-oldenburg.de

### Specialty section:

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

**Received:** 19 May 2017

**Accepted:** 29 September 2017

**Published:** 25 October 2017

### Citation:

Hamann C and Abed Ibrahim L  
(2017) Methods for Identifying  
Specific Language Impairment in  
Bilingual Populations in Germany.  
Front. Commun. 2:16.  
doi: 10.3389/fcomm.2017.00016

This study investigates the performance of 22 monolingual and 54 bilingual children with and without specific language impairment (SLI), in a non-word repetition task (NWRT) and a sentence repetition task (SRT). Both tasks were constructed according to the principles for LITMUS tools (Language Impairment Testing in Multilingual Settings) developed within COST Action IS0804 and incorporated phonological or syntactic structures that are linguistically complex and have been shown to be difficult for children with SLI across languages. For phonology these are in particular (non)words containing consonant clusters. In morphosyntax, complexity has been attributed to factors such as embedding and/or syntactic movement. Tasks focusing on such structures are expected to identify SLI in bilinguals across language combinations. This is notoriously difficult because structures that are problematic for typically developing bilinguals (BiTDs) and monolingual children with SLI (MoSLI) often overlap. We show that the NWRT and the SRT are reliable tools for identification of SLI in bilingual contexts. However, interpretation of the performance of bilingual children depends on background information as provided by parental questionnaires. To evaluate the accuracy of our tasks, we recruited children in ordinary kindergartens or schools and in speech language therapy centers and verified their status with a battery of standardized language tests, assessing bilingual children in both their languages. We consider a bilingual child language impaired if she shows impairments in two language domains in both her languages. For assessment, we used tests normed for monolinguals (with one exception) and adjusted the norms for bilingualism and for language dominance. This procedure established the following groups: 10 typical monolinguals (MoTD), 12 MoSLI, 46 BiTD, and 8 bilingual children with SLI (BiSLI). Our results show that both tasks target relevant structures: monolingual children are classified with 100% accuracy. Crucially, both our tasks distinguish BiTDs from MoSLIs and BiTDs from BiSLIs. The NWRT shows high accuracy and only minimal influence of language dominance. The SRT can be scored as “identical repetition” or as “target structure,” the latter aiming for scoring the mastery of a syntactic structure, ignoring lexical and specific case or gender errors. Focusing on the latter measure, we examine individual cases of BiTDs with unexpected, low scores. We identify first-language dominance as a factor influencing performance but crucially find that testing in the home language in a heritage context might lead to unreliable classifications and that our procedure for determining the clinical group of bilinguals missed cases of selective impairments such as syntactic SLI.

**Keywords:** bilingualism, specific language impairment, sentence repetition, non-word repetition, linguistic complexity

## INTRODUCTION

### Bilingual Language Development and Language Impairment

Recent linguistic research on (specific) language impairment (SLI) has focused on bilingual populations because more and more children grow up bilingually and the challenges of identifying what is typical in bilingual language development and what should be considered an impairment are notorious, see Armon-Lotem et al. (2015) and Marinis et al. (2017) for recent overviews. One such challenge is the finding that SLI may have different manifestations in different languages so that clinical markers widely differ. Extended use of infinitives has been described as a marker of SLI for English (Rice and Wexler, 1996), omission of object clitics for French (Jakubowicz et al., 1998; Paradis et al., 2003) and problems with subject–verb agreement (SVA) together with the use of infinitives and errors in verb placement for German (Clahsen, 1991; Hamann et al., 1998), to mention only some results from well-studied languages. The bigger challenge is, however, that there is an overlap in the linguistic structures that are difficult to master for bilingual children with those structures that are considered clinical markers for SLI in a particular target language; Håkansson and Nettelbladt (1996) were the first to point this out for Swedish, Paradis (2010), Hamann (2012), and Grimm and Schulz (2014) give more recent overviews of similarities and differences. This overlap in error patterns leads to over- and underdiagnosis, see Genesee et al. (2004).

Underdiagnosis occurs if difficulties are ignored based on the argument that delays or deficits in one or both languages often occur in bilingual development, as is the case for bilingual lexical development (Cobo-Lewis et al., 2002; Goldberg et al., 2008; Thordardottir, 2011), the bilingual acquisition of case in German (Schönenberger et al., 2012), or of grammatical gender in Dutch (Cornips and Hulk, 2008). See also Paradis et al. (2016) for a description of long lasting delays in bilingual language development. If, however, such difficulties are taken as evidence for language impairment, overdiagnosis is particularly likely when monolingual norms are applied in tests of the majority language, which might be the weaker language for a child at the time of assessment. Since SLI should be manifest in both languages of a bilingual child, the overlap problem can arguably be avoided if a child's language abilities are assessed in both her languages, the majority language (second language, L2) and the home language (first language, L1). The home language, L1, when spoken most of the time to and by the child in various communicative situations and with various speakers, will be the dominant language before the child is systematically exposed to the L2 in kindergarten or school. This situation often holds for simultaneous, but also for early sequential bilingual children. Even though it has been recommended (Fredman, 2006) that a child be tested in both or, at least, in the dominant language, testing a child in her L1 is often not practicable: there might be no normed tests available for the L1 or the speech language therapist (SLT) cannot administer or evaluate the test in this particular language. In the case of simultaneous bilingual children, it also has to be taken into account that the home language is often a heritage language, i.e., the parents are second or third generation immigrants and speakers of the

language. Heritage situations add further complications: L1 tests, if available, might not be appropriate because the immigrant language might have changed due to contact phenomena as in the case of Immigrant Turkish in Germany (Schroeder and Dollnick, 2013), or, independent of the L1, early acquisition of an L2 might lead to attrition phenomena (Köpke et al., 2004; Montrul, 2008).

The diversity of bilingual profiles and the subtypes of SLI discussed in the literature (Leonard, 1998, 2014) also contribute to the diagnostic difficulties. Bilingual development is crucially influenced by age of onset (AoO), which leads to the definition of simultaneous ( $AoO \leq 3$ ) bilingualism, early ( $3 < AoO < 4$ ) and late ( $AoO \geq 4$ ) sequential child bilingualism (also called child L2), as well as to a clear distinction of child and adult L2 speakers, see Meisel (2009) for a discussion of early and late child L2.<sup>1</sup> Length of exposure (LoE), quantity and quality of input, and socioeconomic status (SES) also contribute crucially to bilingual language development so that background information about these factors is essential for the assessment of language samples and the interpretation of test results. Though SLI frequently concerns both phonological and morphosyntactic development, selective impairments have been identified, such as grammatical/syntactic (van der Lely, 1998) or semantic SLI (Schulz and Roeper, 2011), see also Friedmann and Novogrodsky (2011). The diversity of subtypes of SLI contributes to the problems in identifying language impairment in bilingual children.

### The Language Impairment Testing in Multilingual Settings (LITMUS) Tools for Crosslinguistic Research

Given these difficulties, several approaches can be explored. First, existing assessment tools can be normed for bilingual populations. Second, existing tools normed for monolinguals can be applied adjusting the norms for bilingualism and according to the status of the language being tested as the dominant or weaker language, see the recommendations by Thordardottir (2015) described in Section “Participants and Procedure for Verification of Clinical Status.” Third, new tools can be constructed according to linguistic principles that allow crosslinguistic application, such as the tools developed during and following COST Action IS0804. These are called LITMUS tools and are described in Armon-Lotem et al. (2015). Of specific interest here are the LITMUS principles outlined in Chiat (2015) for non-word repetition tasks (NWRs) and by Marinis and Armon-Lotem (2015) for sentence repetition tasks (SRTs) and the Questionnaire for Parents of Bilingual Children (PaBiQ) described by Tuller (2015). These three tasks were central in a French–German joint project (BiLaD – bilingual language development)<sup>2</sup> investigating monolingual and bilingual children with and without language impairment and with Arabic,

<sup>1</sup>Note that authors often use their own definitions, e.g., Schulz and Tracy (2011) define children with an AoO < 24 months as simultaneous bilinguals.

<sup>2</sup>The project was funded by DFG (German Science Foundation) grants HA 2335/6-1, RO 923/3-1, CH 1112/2-1 to Cornelia Hamann, Monika Rothweiler, and Solveig Chilla as well as an ANR (French Science Agency) to Laurice Tuller as principal investigator.

Portuguese and Turkish as home languages,<sup>3</sup> of which we report the German data here.

We focus on non-word repetition and sentence repetition since such tasks have been shown to reliably identify SLI in monolinguals (Conti-Ramsden et al., 2001) and are often part of standard assessment tools. Such tests usually assess working memory (WM), see Archibald and Gathercole (2006) but can be constructed so that they measure the command of phonological or syntactic representations/derivations (see Gallon et al., 2007 for non-word repetition; Polišenská et al., 2014 for sentence repetition). In SRTs, this can be achieved by taxing memory with number of words and vocabulary so that a successful parse of the sentence is a necessary condition for successful repetition. In addition, structurally minimal pairs should be incorporated to identify the locus of difficulty: in the case of embedding, a finite complement clause can be contrasted with a coordination structure, which also contains two propositions but does not embed one into the other. The LITMUS tasks incorporate linguistically complex (syntactic or phonological) structures and operations known to be difficult for children with SLI crosslinguistically or in a particular language, such as SVA or topicalization in German. For syntax, especially structures involving syntactic movement, particularly Wh-movement, i.e., fronted interrogative or relative pronouns (see Hamann et al., 1998; van der Lely, 1998; Friedmann and Novogrodsky, 2011), as well as embedding (Hamann and Tuller, 2014) have been crosslinguistically identified by recent research as vulnerable in children with SLI. A particular difficulty has been identified for structures that involve movement and contain intervening elements between the source of the moved element and its landing site (Rizzi, 2004; Friedmann et al., 2015). The latter difficulty occurs in object Which-questions and in object relative clauses containing a lexical subject. In contrast to the difficulties encountered by children with SLI, a typically developing bilingual child might have problems with vocabulary or grammatical features that do not have semantic content (uninterpretable linguistic features, such as number agreement on the verb, Tsimpli and Dimitrakopoulou, 2007) and might even avoid complexity, but should in principle not be overtaxed by structures involving movement or embedding. Recent results indicate that SRTs incorporating structures involving these operations can be successfully applied in bilingual settings for identifying SLI, see Marinis and Armon-Lotem (2015), Tuller et al. (2015), and Fleckstein et al. (2016). As to non-word repetition and phonological complexity, recent studies show that syllables containing branching onsets or a coda are particularly difficult for children with SLI, but are mastered by typical bilinguals (Marshall and van der Lely, 2009; Ferré et al., 2012; dos Santos and Ferré, 2016; Grimm and Hübner, in press). NWRTs can be constructed to incorporate quasi-universal non-words or non-words conforming to phonotactic and/or morphophonological constraints of a specific language. Especially the quasi-universal type can be used successfully with bilingual children after only a short time

of exposure to the target language, independent of SES and L2 experience, see Chiat and Polišenská (2016). Thordardottir and Brandeker (2013) compared performance of bilingual children in an NWRT and an SRT to performance on receptive vocabulary and found the latter more affected by levels of previous exposure than NWRT and SRT, with NWRT and SRT showing acceptable sensitivity levels. Quite recently, LITMUS NWRTs and SRTs have been studied as to their diagnostic accuracy in bilingual populations. Boerma et al. (2015) use a quasi-universal LITMUS NWRT and report excellent accuracy for their population of bilingual children with Dutch as L2. Armon-Lotem and Meir (2016) find good accuracy for their Hebrew LITMUS SRT in Russian-Hebrew bilingual children, whereas the accuracy for their NWRT, with word-like items incorporated, is described as fair. The arguably good diagnostic accuracy of NWRT and SRT in monolingual and bilingual populations (but see also Gutiérrez-Clellen and Simon-Cereidjido, 2010) led us to develop and investigate an SRT for German and to adopt the NWRT developed by Grimm et al. (2014) and investigate it with our bilingual population.

## Research Questions and Aims of the Present Study

This study presents data from 54 bilingual children living in Germany with Arabic, Portuguese and Turkish as their home language, comparing them to 22 monolingual children. The overall aim of the study is to investigate two new LITMUS tools for German, a sentence repetition and a non-word repetition task developed according to the LITMUS principles (COST Action IS0804, Chiat, 2015; Marinis and Armon-Lotem, 2015). We want to know in particular whether they are able to identify SLI in bilinguals. For the NWRT, we want to know how accurate it is for our population, and we specifically investigate the German SRT as a new method and discuss its evaluation by different scoring procedures. As a first step, we therefore investigate the performance of monolingual children with and without SLI on these tasks. For evaluating the accuracy of the new tasks in bilingual children, groups of typically developing bilingual children and of bilingual children with SLI were defined. For this goal, mono- and bilingual children without any history of language problems were recruited in ordinary kindergartens and schools and children with a diagnosis of SLI (mono- and bilingual) were recruited in speech language centers or private practice. This initial grouping was verified and if necessary corrected by using norm-referenced L1 and L2 tests adjusting the norms as suggested by Thordardottir (2015) and described in more detail in Section “Participants and Procedure for Verification of Clinical Status.” This procedure, as pointed out by Thordardottir (2015), is not unproblematic and will be discussed with respect to the status of the home languages as heritage languages and the different subgroups of SLI. We will then proceed to show that tests in the L2 can be very reliable, especially the LITMUS tasks. It will also emerge, however, that in most cases a combination of tests should be applied to achieve good diagnostic accuracy.

<sup>3</sup>See Fleckstein et al. (2016), Almeida et al. (2017), and dos Santos and Ferré (2016) for results on the French versions of the NWRT and the SRT.

## METHODS AND PROCEDURES

### Participants and Procedure for Verification of Clinical Status

We investigated bilingual children with Arabic, Portuguese and Turkish as home languages. These languages were chosen because there are substantial groups of Arabic, Portuguese and Turkish immigrants in Germany<sup>4</sup> and because the language communities differ from each other, so that comparisons can be made. Children were recruited in kindergartens, schools and in speech language therapy centers. The study was carried out in accordance with the compliance form, transaction number 20120416505890730506, of the German Science Foundation and the recommendation of the “Kommission für Forschungsfolgenabschätzung und Ethik” (Commission for the Evaluation of Research Consequences and Ethics) of the Carl-von-Ossietzky University of Oldenburg (rf. Drs. 21/16/2013). Written informed consent was obtained from all adult research participants as well as from the parents/legal guardians of all minors. Written informed consent was obtained from the parents both for the purposes of data collection through the Parental Questionnaire as well as for the purposes of their children's participation in this research. The protocol was approved by the “Kommission für Forschungsfolgenabschätzung und Ethik” of the Carl-von-Ossietzky University of Oldenburg.

The age range of the children was chosen as 5;5–9;4 years since this includes the last year of kindergarten and the crucial first 2 or 3 years in primary school. We recruited 22 monolingual children, 10 typically developing and 12 with a diagnosis of SLI. In addition, 38 typically developing bilingual children were selected in Germany as well as 16 bilingual children in SLT, see (Table 1). We included only bilingual children with an LoE of more than 24 months. Our group includes simultaneous and sequential bilinguals, where we define the latter as children who were systematically exposed to their L2 at the age of 36 months or later.

The status of all of these children as typical or language impaired was then verified by a battery of tools following part of the protocol suggested by Thordardottir (2012). We first tested for non-verbal cognition with the German version of Raven's colored progressive matrices (CPM), see Bulheller and Häcker (2002), excluding children who scored below percentile 9 (the cutoff for low-average

non-verbal intelligence, equivalent to an IQ-score  $\leq 80$  according to Wechsler's IQ scale). We also collected a narrative language sample in each of a child's languages. For the latter, we used the materials provided by the Multilingual Assessment Instrument for Narratives (MAIN), another LITMUS task perfected within Cost Action IS0804 (Gagarina et al., 2015), but did not evaluate the narratives according to the MAIN protocol. Instead, we used the material to (a) judge the expressive abilities of a child in each of her languages to confirm or disconfirm the status of a language as the weaker or the dominant language and (b) to scan the material for clinical markers of SLI such as SVA errors in German.<sup>5</sup> In our sample no child was excluded because of performance in CPM and all children in the bilingual groups had at least receptive command of two languages.

Following many researchers on SLI, see Leonard (2014), Tomblin et al. (1997), and also Thordardottir (2015), we classified a monolingual child as having SLI (MoSLI) whenever performance was below  $-1.25$  SDs in two language domains in appropriate norm-referenced tests. Relevant language domains in this context are phonology (receptive and productive), receptive and productive vocabulary and comprehension and production of morphosyntax. For bilingual children, we followed Thordardottir (2015), who suggests the following norm adjustments for norm-referenced tests with monolingual norms: A bilingual child is considered SLI if she scores  $-1.5$  SDs below mean scores of typical monolingual peers in her dominant language,  $-2.25$  SD in her weaker language, and  $-1.75$  SD in either language if she is a balanced bilingual. We are aware that these cutoffs were calculated for groups of simultaneous bilingual children.

We administered three norm-referenced L2 tests, the LiSe-DaZ, the WWT and the PLAKSS-II, covering morphosyntax, lexicon and phonology separately, and the ELO-L for Arabic, the PALPA-P and GOL-E for Portuguese, and the TEDIL for Turkish as L1 tests, see Section “Standardized L2 and L1 Tests” for details. The results were interpreted on the background information provided by the PaBiQ. In particular, the calculation of children's language dominance allowed the application of adjusted cutoffs. With the help of these adjustments for tests providing monolingual norms, we classified a bilingual child as language impaired only if the child performed below the respective cutoffs in two language domains in both of her languages. For the TEDIL, because it provides only two composite values, we used the suggested monolingual norm of  $-1.0$  adjusting it according to dominance. For the LiSe-DaZ, which provides bilingual norms for sequential bilingual children (defined by the authors as  $\text{AoO} > 2$ ) and also monolingual norms, we used a cutoff of  $-1.25$  SD. Since expressive vocabulary is a notorious domain of difficulty for bilingual children in both languages, we decided to count the lexicon as a single domain and consider a bilingual child as typically developing in her lexicon if she scored above the appropriate cutoffs in receptive vocabulary. This leads to the classification of participants as shown in Table 2, which also shows our control groups, the monolingual children with and without SLI.

<sup>4</sup>This also holds for France, which makes cross-country comparisons possible in the project.

**TABLE 1** | Bilingual children in Germany: children not in speech language therapy (SLT) and children in SLT.

|                          | Not in SLT   | In SLT        |
|--------------------------|--------------|---------------|
| Bilingual: L1 Arabic     | 7            | 4             |
| Bilingual: L1 Portuguese | 19           | 1             |
| Bilingual: L1 Turkish    | 12           | 11            |
| Total no. participants   | 38           | 16            |
| M age (SD)               | 82.5 (12.80) | 84.68 (14.96) |
| Range                    | 66–107       | 65–109        |
| Gender                   | 21 F, 17 M   | 6 F, 10 M     |

<sup>5</sup>Further evaluation of narrative micro- and macro-structure according to the MAIN protocol will be the next step.

**TABLE 2** | Participants including monolingual children and final status of bilingual children as BiTD and BiSLI: age at testing (months), colored progressive matrices (CPM) scores (percentile ranks), and gender.

|                      | MoTD ( <i>n</i> = 10)      | MoSLI ( <i>n</i> = 12)  | BiTD                    |                         |                         | BiSLI <sup>a</sup> ( <i>n</i> = 8) |                        |
|----------------------|----------------------------|-------------------------|-------------------------|-------------------------|-------------------------|------------------------------------|------------------------|
|                      |                            |                         | BiTD-A ( <i>n</i> = 8)  | BiTD-P ( <i>n</i> = 19) | BiTD-T ( <i>n</i> = 19) | Total ( <i>n</i> = 46)             |                        |
| <b>Simult./total</b> |                            |                         |                         |                         |                         | <b>40/46</b>                       | <b>8/8</b>             |
|                      | <b>Mean (SD) and range</b> |                         |                         |                         |                         |                                    |                        |
| Age at testing       | 75.90 (8.99)<br>66–92      | 81.75 (13.41)<br>68–112 | 88.50 (12.11)<br>70–107 | 80.95 (13.51)<br>66–106 | 88.68 (14.32)<br>70–111 | 85.45 (13.87)<br>66–111            | 82.12 (16.4)<br>65–109 |
| CPM                  | 81.20 (13.98)<br>56–100    | 53.41 (23.39)<br>25–99  | 47.37 (30.91)<br>9–93   | 71.63 (20.59)<br>28–94  | 67.57 (22.39)<br>36–100 | 65.73 (24.39)<br>9–100             | 59.12 (26.58)<br>27–98 |
| Gender               | 8 F, 2 M                   | 5 F, 7 M                | 5 F, 3 M                | 9 F, 10 M               | 10 F, 9 M               | 25 F, 21 M                         | 2 F, 6 M               |

<sup>a</sup>BiSLI group (*n* = 8): 3 L1 Arabic, 1 L1 Portuguese, and 4 L1 Turkish.

Comparing the initial groups from **Table 1** to the classification achieved by L1 and L2 testing in **Table 2**, it is striking that the bilingual population with language impairment has been cut in half. Our procedure, and testing in L1 in particular, has uncovered eight potential cases of overdiagnosis.

The four final groups (MoTD, MoSLI, BiTD, and BiSLI) were comparable concerning non-language variables such as age, non-verbal intelligence, and SES (see **Table 2**).<sup>6</sup> A Kruskal–Wallis non-parametric test<sup>7</sup> revealed no significant differences in terms of age at testing between the four groups of participants [ $\chi^2(3, N = 76) = 4.061, p = 0.255$ ]. The age difference remains statistically insignificant even when the BiTD group is split by the children's home language into three subgroups (BiTD-A, BiTD-P, and BiTD-T) [ $\chi^2(5, N = 76) = 7.782, p = 0.169$ ]. Although the Kruskal–Wallis test revealed a marginally significant difference with respect to the four groups' non-verbal intelligence [ $\chi^2(3, N = 76) = 7.689, p = 0.053$ ], *post hoc* Mann–Whitney *U* test applying Bonferroni correction revealed only one significant comparison between the MoTD and MoSLI group ( $U = 154, p = 0.036, r = 0.348$ ). Nevertheless, all of the children in the MoSLI group have normal non-verbal intelligence. We further checked whether the L1 Arabic, L1 Portuguese, and L1 Turkish typically developing children were comparable for SES as measured by years of mother's education. Since no significant differences were observed [ $\chi^2(2, N = 46) = 0.181, p = 0.913$ ], the three subgroups were collapsed into one BiTD group. A Kruskal–Wallis test also revealed that the BiTD and BiSLI groups were similar with respect to SES.

## Standardized L2 and L1 Tests

For L1 and L2 assessment, we chose standardized tests in both languages that are commonly used in speech language therapy and are normed for the age range investigated here—or for which norms can be extended, see **Table 3** for an overview. An important decision for assessment in German was made in the choice of the LiSe-DaZ (Schulz and Tracy, 2011), which is the first German standardized test normed not only for

monolinguals but also for sequential bilingual children between 3;0–7;11. Comprehension of negation, of constituent questions, and of telic events is tested. The assessment of production targets SVA, sentence complexity, case marking, and word classes (prepositions, main verbs, auxiliaries, focus particles, and conjunctions). All subtasks except those for sentence complexity and SVA provide *t* values. The recommendation of the authors is to consider a child “at risk for language impairment if she performs more than 1 SD below *t* = 50 in two of the 9 subtests with *t* values” (Grimm and Schulz, 2014, p. 831). This procedure excludes an area of morphosyntax, SVA, which has been discussed as clinical marker for (bilingual) SLI in German (Rothweiler et al., 2012), and does not allow separate evaluation of performance in production and comprehension. We departed from the authors' own rating procedure by (a) setting the cutoff at  $-1.25$  SD and (b) ignoring the results of the case task (see Lein et al., 2016; Abed Ibrahim et al., in press). The test does not offer norms for simultaneous bilingual children with an AoO < 24 months or bilingual children older than 8 years. For older children, however, a cutoff of  $-0.5$  SD is suggested by the authors, and for simultaneous bilinguals monolingual norms can be applied whenever German is the dominant language. Since the LiSe-DaZ is an assessment of comprehension and production of morphosyntax only, other domains of language had to be evaluated with separate tests. We chose the WWT (Glück, 2007) for evaluation of lexical reception and production and the PLAKSS-II (Fox-Boyer, 2014) for evaluation of phonology. For classifying a child as BiSLI it was necessary that she performs below adjusted cutoffs in two domains of L1 and two domains of L2. For the L2 tests, this implies that she had to perform below cutoffs in two subtasks of the LiSe-DaZ (morphosyntax) combined with low performance in either the PLAKSS-II (phonology) or the receptive subtest of the WWT (vocabulary), or she had to perform below cutoffs in the receptive part of the WWT and in the PLAKSS-II.

Turning to the three different L1s, we chose the ELO-L for Arabic (Zebib et al., 2017). It uses word repetition for phonological abilities, picture naming and picture selection for lexical production and reception, sentence completion and picture selection for assessing morphosyntax. It exists in two versions, for younger (3;0–5;11) and older (6;0–7;11) children, is normed for both versions on a large and mixed population, and takes 30–45 min to

<sup>6</sup>Information on SES is only available for bilingual children in our data set.

<sup>7</sup>See Section “Data Analysis” for the choice of statistical tests, taking account of the unequal group sizes.

**TABLE 3** | Standardized tests used for language assessment in Arabic, German, Portuguese, and Turkish: overview.

| Language | Name of test           | Language skill assessed |                      |                       |                                 |                                               | Scoring method                                       | Norming population                               |
|----------|------------------------|-------------------------|----------------------|-----------------------|---------------------------------|-----------------------------------------------|------------------------------------------------------|--------------------------------------------------|
|          |                        | Phonology               | Receptive vocabulary | Expressive vocabulary | Morphosyntax comprehension      | Morphosyntax production                       |                                                      |                                                  |
| Arabic   | ELO-L <sup>a</sup>     | Word repetition         | Picture selection    | Picture naming        | Picture–sentence matching       | Sentence completion                           | Individual subtest scores and global score           | 3;0–7;11                                         |
| German   | WWT 6–10 <sup>b</sup>  | –                       | Picture selection    | Picture naming        | –                               | –                                             | Individual subtest scores                            | 5;6–10;11                                        |
|          | LiSe-DaZ <sup>c</sup>  | –                       | –                    | –                     | Picture–sentence matching, TVJT | Story, sentence completion, lead-in questions | Individual subtest scores                            | Monolinguals: 3;0–6;11<br>Bilinguals: 3;0–7;11   |
|          | PLAKSS-II <sup>d</sup> | Picture naming          | –                    | –                     | –                               | –                                             | Individual subtest scores                            | 2;6–7;11                                         |
| EP       | PALPA-P <sup>e</sup>   | Non-word repetition     | Picture selection    | Picture naming        | Picture selection               | Sentence repetition                           | Individual subtest scores                            | 5;0–9;11 (with some missing norms for all tasks) |
|          | GOL-E <sup>f</sup>     | –                       | Word definition      | Antonyms naming       | –                               | Sentence construction                         | Individual subtest scores and global score           | 5;07–10;00                                       |
| Turkish  | TEDIL <sup>g</sup>     | –                       | Picture selection    | Picture naming        | Picture selection               | Sentence completion                           | 2 Composite scores, 1 production and 1 comprehension | 2;0–7;11                                         |

<sup>a</sup>Zebib et al. (2017).<sup>b</sup>Glück (2007).<sup>c</sup>Schulz and Tracy (2011).<sup>d</sup>Fox-Boyer (2014).<sup>e</sup>Castro et al. (2007).<sup>f</sup>Sua-Kay and Santos (2014).<sup>g</sup>Topbaş and Güven (2013).

EP, European Portuguese; TVJT, truth value judgment task.

administer. The test takes into account the bilingual situation in Lebanon and was translated by native speakers to other varieties of Arabic such as Algerian, Egyptian, Moroccan, Tunisian, Libyan, Palestinian, and recently Syrian.

The PALPA-P (Provas de Avaliação da Linguagem e da Afasia em Português) was adapted by Castro et al. (2007) from the *Psycholinguistic Assessments of Language Processing in Aphasia* by Kay et al. (1996) and provides a linguistically well-controlled instrument for the assessment of children with European Portuguese as L1. The test evaluates the domains of phonology, lexical production and reception as well as morphosyntactic production and comprehension. It is normed for children aged 5;0–9;0 (with certain gaps, especially in the lexical evaluation) and takes about 50 min to administer. Scoring is correct (1) or incorrect (0). Since there are age gaps in the norming population for the lexical tasks in the PALPA-P, we used the GOL-E (Sua-Kay and Santos, 2014) for lexical production and comprehension with norms for children between 5;7 and 10;0 years of age.

For Turkish, we chose the TEDIL by Topbaş and Güven (2013), an adaptation of the TELD-3, which has been normed for children aged 2;0–7;11. It exists in two different versions for younger and older children, and measures comprehension and production in morphosyntax, morphology and lexical semantics. The task does not specifically test for phonology, but has a subtask for lexical reception and two further receptive tasks on lexical relations. For morphosyntax, there is a comprehension

and a production part in the form of a repetition task. Norms exist for composite scores of reception and expression only, not for individual subscores.

## The LITMUS-PaBiQ

An important assessment tool for the evaluation of language abilities in bilingual children is a questionnaire that can provide the background for the interpretation of test results. Information about the child's language exposure and use, current and in her early years of development, is essential and allows determination of language dominance. For this purpose, we chose the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015), which was developed within COST Action IS0804 based on questionnaires developed in Paradis et al. (2010) and Paradis (2011). We used a German translation of the questionnaire as well as translations into Arabic, Portuguese, or Turkish so that parents could choose in which language the interview, by phone or in person, would be conducted.

Parental questionnaires, and the PaBiQ in particular, pay special attention to age of first systematic language exposure (AoO), LoE, quality and quantity of input at home, and other everyday situations and also provide information about parents' education, which can be taken as an indication of SES. Apart from these variables known to impact bilingual development, indicators for language impairment were also incorporated into the questionnaire. These include early language development (first words and first sentences) and family history of language difficulties. The

latter variables allow calculating a No-Risk Index, a reliable indicator for the French group of children investigated in the BiLaD project (Almeida et al., 2017) and currently under investigation for the whole group and the German bilinguals in particular.

Returning to the factors influencing bilingual development, they allowed us to determine an L2 Exposure Index and an L1 Exposure Index. These indexes were calculated by weighing factors such as AoO, LoE, language use, and richness at home, at school, in extracurricular activities, before and after the age of 4 years. The Language Dominance Index (LDI) can be calculated as the difference between the L2 and the L1 exposure indexes. Given the individual contributions of the factors in L1 and L2, the LDI ranges from  $-50$  to  $+50$ . For the project, several cutoff points were explored and compared with impressions of bilingual investigators, specifically taking into account free conversation and the samples of spontaneous speech collected for each child in each language (see also Almeida et al., 2017). Following that procedure, we define bilingual children in Germany as balanced if they score between the values of  $-5$  and  $+5$  of the LDI ( $-5 \leq \text{LDI} \leq +5$ ). Children with an LDI below  $-5$  are considered L1 dominant whereas children with an LDI above  $+5$  are classified as L2 dominant.

## The New German LITMUS Repetition Tasks

### The German LITMUS NWRT

Since the goal is to not disadvantage bilingual children when assessing their phonological abilities, the NWRT (see Grimm and Hübner, in press) was designed to include vowels and consonants common in most languages of the world, at the same time targeting complex phonological structures, i.e., consonant clusters, known to cause difficulty in children with SLI, see Chiat (2015) and Ferré et al. (2012). In particular, the NWRT contained a language-independent (LI) part and a language-dependent (LD), see Grimm and Hübner (in press)<sup>8</sup> and Abed Ibrahim and Hamann (2017) for a detailed description of the task. There were maximally three syllables in the non-words so that memory effects would only minimally influence performance. The 30 non-words of the LI part were built using phonemes and phonotactic properties well attested crosslinguistically (Maddieson et al., 2011). Differing from the universal NWRT discussed in Chiat and Polišenská (2016), the task does not only contain simple CV syllables but also syllables with branching onsets of the type “CCV” and a final consonant coda (coda, CVC#), which are nonetheless characterized by their crosslinguistic frequency (Maddieson, 2006). We expect monolingual and bilingual children with SLI to have difficulties with these phonologically complex structures whereas typical monolingual and bilingual children should master them. The LD part contains 36 non-words with two more additional German consonants /s, ʃ/ and more syllable types as shown in Table 4. Since sC# and #Cs sequences are not unique to German but violate the Sonority Sequencing Principle, they are difficult for

**TABLE 4 |** Overview of the German LITMUS non-word repetition task.

|                                   | Vowels    | Consonants   | Syllable types           | Examples                |
|-----------------------------------|-----------|--------------|--------------------------|-------------------------|
| Language-independent part         | /a, i, u/ | /p, k, f, l/ | CV<br>CCV<br>CVC#        | pilu<br>flaplu<br>kafip |
| 30 items                          |           |              |                          |                         |
| 23 test items                     |           |              |                          |                         |
| 7 control items (e.g., pilu, paf) |           |              |                          |                         |
| Language-dependent part           | /a, i, u/ | /p, k, f, l/ | same syllable types plus |                         |
| 36 items                          |           | plus         | #sCV                     | skifapu                 |
| 32 test items                     |           | /s/          | #sCCV                    | ʃplaklu                 |
| 4 control items (e.g., kas, ʃaku) |           | /ʃ/          | Cs#                      | fikapuks                |
|                                   |           |              | internal/s/              | Kufiski                 |

children with SLI (dos Santos and Ferré, 2016) but should not be problematic for typically developing children.

The task, in the form of a PowerPoint presentation (PPT), is easy to administer and takes about 5–10 min. It is appealing to children since they are told that it is an alien who is trying to teach them his language. Items were presented in pseudo-randomized order through headphones. Scoring took into account whole item accuracy, disregarding systematic substitutions, e.g., /t/ for /k/, as well as errors in minimally different vowels or voicing of consonants. Following Grimm and Hübner (in press), we also disregarded substitution of extrametrical /ʃ/by [s] since their substitution does not lead to a phonemic contrast in syllable initial position in German.

### The German LITMUS SRT

The German SRT, first introduced by Hamann et al. (2013), was constructed in parallel to the French task (Fleckstein et al., 2016; Almeida et al., 2017) during COST Action IS0408 incorporating the LITMUS principles (Marinis and Armon-Lotem, 2015). It thus contains complex structures known to be difficult for children with SLI crosslinguistically, including object questions, subject and object relative clauses, finite complement clauses and passives, as well as structures identified as milestones in the acquisition of German word-order properties such as topicalization, and the sentence bracket, see examples (5) and (1). See Hamann et al. (2017) and Lein et al. (2016) for details on the German SRT and Hamann (2015) for an overview of SLI in German.

The version<sup>9</sup> of the German LITMUS SRT investigated in this study contains 45 sentences with three levels of increasing complexity controlled for number of syllables in each level (five conditions per level and three items per condition). Stimuli are presented in randomized order *via* a child friendly PPT. The levels arise through adding factors of complexity such as Wh-movement, embedding, intervention and the fact that two propositions are presented. Thus level 1 contains simple declaratives and assesses SVA, tense and the sentence bracket, (1). Level 2 includes object

<sup>8</sup>We particularly thank Angela Grimm for sharing the task with us.

<sup>9</sup>The original long version of the German LITMUS-SRT was shortened to meet the needs of the age range investigated in the BiLaD project.

questions with an intervening lexical NP subject. Following Rizzi (2004), these are Which-NP questions, where the interrogative constituent contains a lexical NP as restriction, which has moved over a lexical subject as in “Welchen *Clown* umarmt der *Wiking* <welchen *Clown*>—which *clown* does the *viking* hug <which *clown*>.” These are contrasted with questions where the question constituent does not carry a lexical restriction (*wen-whom*, bare Wh) and therefore there is no intervention. All questions ask for masculine persons with unambiguous case marking, see (2a) and (2b). The task also contains finite (3), and non-finite complement clauses contrasting with coordinate structures. Level 3 contains long passives, subject relatives, object relatives with, (4), and without a lexical intervener, as well as topicalizations (5). **Table 5** gives a summary.

#### (1) Sentence bracket:

|                         |                |                   |                         |                  |                            |
|-------------------------|----------------|-------------------|-------------------------|------------------|----------------------------|
| Die<br>The/ <i>nom.</i> | Köchin<br>cook | <b>hat</b><br>has | den<br>the/ <i>acc.</i> | Cowboy<br>cowboy | <b>geweckt</b><br>woken up |
|-------------------------|----------------|-------------------|-------------------------|------------------|----------------------------|

“The cook woke the cowboy up”

#### (2a) Bare WH

|                                |                    |                  |                    |                  |
|--------------------------------|--------------------|------------------|--------------------|------------------|
| <b>Wen</b><br>Who/ <i>acc.</i> | umarmt der<br>hugs | the/ <i>nom.</i> | Pinguin<br>penguin | heute?<br>today? |
|--------------------------------|--------------------|------------------|--------------------|------------------|

“Whom does the penguin hug today?”

#### (2b) Which-NP

|                                      |                       |                   |                                             |
|--------------------------------------|-----------------------|-------------------|---------------------------------------------|
| <b>Welchen</b><br>Which/ <i>acc.</i> | <b>Clown</b><br>Clown | besucht<br>visits | der Zauberer?<br>the/ <i>nom.</i> Magician? |
|--------------------------------------|-----------------------|-------------------|---------------------------------------------|

“Which clown does the magician visit?”

#### (3) Finite complement clause:

|                         |                 |                 |                     |                         |                  |                         |                  |                      |
|-------------------------|-----------------|-----------------|---------------------|-------------------------|------------------|-------------------------|------------------|----------------------|
| Der<br>The/ <i>nom.</i> | Prinz<br>Prince | will,<br>wants, | <b>dass</b><br>that | der<br>the/ <i>nom.</i> | Ritter<br>knight | die<br>the/ <i>acc.</i> | Affen<br>monkeys | <b>jagt</b><br>hunts |
|-------------------------|-----------------|-----------------|---------------------|-------------------------|------------------|-------------------------|------------------|----------------------|

“the prince wants that the knight hunts the monkeys”

#### (4) Object relative with intervention:

|          |             |                         |                 |                                |                                |                         |                |
|----------|-------------|-------------------------|-----------------|--------------------------------|--------------------------------|-------------------------|----------------|
| Ich<br>I | sehe<br>see | den<br>the/ <i>acc.</i> | Clown,<br>Clown | <b>den</b><br>who/ <i>acc.</i> | <b>der</b><br>the/ <i>nom.</i> | <b>Wiking</b><br>viking | umarmt<br>hugs |
|----------|-------------|-------------------------|-----------------|--------------------------------|--------------------------------|-------------------------|----------------|

“I see the clown who(m) the viking hugs”

**TABLE 5** | Overview of the German LITMUS sentence repetition task.

| Level 1                     | Level 2                        | Level 3                                         |
|-----------------------------|--------------------------------|-------------------------------------------------|
| SVO-present                 | Object Who-question (bareWH)   | Passive                                         |
| SVO-simple past             | Object Which-question (Wh-NP)  | Topicalization                                  |
| Sentence bracket (aux)      | Coordination (Coord)           | Subject relatives (SR)                          |
| Sentence bracket (particle) | Non-finite complement Cl       | Object relatives without intervener (OR –intv.) |
| “Werden” control            | Finite complement Cl (CompFin) | Object relatives with intervener (OR +intv.)    |

#### (5) Topicalization

|                                |                     |                   |                         |                      |                     |
|--------------------------------|---------------------|-------------------|-------------------------|----------------------|---------------------|
| <b>Den</b><br>The/ <i>acc.</i> | <b>Koch</b><br>cook | besucht<br>visits | der<br>the/ <i>nom.</i> | Zauberer<br>magician | als ersten<br>first |
|--------------------------------|---------------------|-------------------|-------------------------|----------------------|---------------------|

“The cook, the magician visits first.”

The task takes about 10 min to administer. Items are scored as 0/1 using different criteria for this rating. “Identical repetition” only disregards phonological errors and is the fastest and easiest way of scoring. Since lexical substitutions and omissions are counted as errors in this scoring method, difficulties that bilingual children have with vocabulary will clearly show in this measure. An alternative method is “target structure” which aims to ascertain that a child masters certain complex structures in principle. It compensates for L2 errors such as lexical substitutions and systematic recurrent case errors as well as gender errors that do not affect the realization of the targeted structure see the examples in (6) to (8). Errors not affecting the realization of the target structure in the examples are given in bold print:

#### (6) Target structure: (sentence bracket)

|                         |                |            |                         |                  |                     |
|-------------------------|----------------|------------|-------------------------|------------------|---------------------|
| Die<br>The/ <i>nom.</i> | Köchin<br>cook | hat<br>has | den<br>the/ <i>acc.</i> | Cowboy<br>cowboy | geweckt<br>woken up |
|-------------------------|----------------|------------|-------------------------|------------------|---------------------|

“The cook woke the cowboy up”

#### Child repetition:

|                         |                |            |                         |                              |                     |
|-------------------------|----------------|------------|-------------------------|------------------------------|---------------------|
| Die<br>The/ <i>nom.</i> | Köchin<br>cook | hat<br>has | den<br>the/ <i>acc.</i> | <b>Clown</b><br><b>clown</b> | geweckt<br>woken up |
|-------------------------|----------------|------------|-------------------------|------------------------------|---------------------|

lexical substitution error, target structure score: 1

#### (7) Target structure: (long passive)

|                         |                |            |           |                         |                |                  |                     |
|-------------------------|----------------|------------|-----------|-------------------------|----------------|------------------|---------------------|
| Die<br>The/ <i>nom.</i> | Oma<br>grandma | wird<br>is | von<br>by | dem<br>the/ <i>dat.</i> | großen<br>tall | Cowboy<br>cowboy | geärgert<br>annoyed |
|-------------------------|----------------|------------|-----------|-------------------------|----------------|------------------|---------------------|

“The grandma is annoyed by the tall cowboy”

#### Child repetition:

|                         |                |            |           |                                          |                |                  |                     |
|-------------------------|----------------|------------|-----------|------------------------------------------|----------------|------------------|---------------------|
| Die<br>The/ <i>nom.</i> | Oma<br>grandma | wird<br>is | von<br>by | <b>den</b><br><b>the/</b><br><b>acc.</b> | großen<br>Tall | Cowboy<br>cowboy | geärgert<br>annoyed |
|-------------------------|----------------|------------|-----------|------------------------------------------|----------------|------------------|---------------------|

systematic case error in passives, *von dem* rendered as *von den*, target structure score:1

#### (8) Target structure: (SVA, third, sg)

|                         |                 |             |                  |                              |                      |
|-------------------------|-----------------|-------------|------------------|------------------------------|----------------------|
| Der<br>The/ <i>nom.</i> | kleine<br>small | Hund<br>dog | bringt<br>brings | die<br>the/ <i>fem.-acc.</i> | Zeitung<br>newspaper |
|-------------------------|-----------------|-------------|------------------|------------------------------|----------------------|

“The small dog brings the newspaper”

#### Child repetition:

|                         |                 |             |                  |                                     |                       |
|-------------------------|-----------------|-------------|------------------|-------------------------------------|-----------------------|
| Der<br>The/ <i>nom.</i> | kleine<br>small | Hund<br>dog | bringt<br>brings | <b>den</b><br><b>the/masc.-acc.</b> | Zeitung.<br>newspaper |
|-------------------------|-----------------|-------------|------------------|-------------------------------------|-----------------------|

gender error, target structure score: 1

Using this method might miss measuring the total effect of linguistic complexity. Quite often, several errors occur in complex structures, not necessarily however on the specific marker of the structure itself. To give an example: When a finite complement clause is targeted, the structural difficulty might be manifest in an omission of the complementizer (*dass—that*) and a simple juxtaposition of clauses. This would clearly be 0 for scoring as “target structure.” The difficulty could surface in lexical substitution, however, or there could be additional errors unrelated to the complementizer. Since “target structure” is a measure that does not penalize bilingual children and can establish whether structures such as finite complement clauses are acquired or not, we nevertheless use this measure for scoring German SRTs in addition to the measure of identical repetition.

## Research Questions Concerning the German LITMUS Repetition Tasks

As stated in Section “Research Questions and Aims of the Present Study,” we want to know whether the German LITMUS SRT and NWRT are able to identify language impairment in bilingual settings. For this purpose, we first ask whether the tasks successfully identify SLI in monolingual German children. We also want to know in how far our tasks can be used as a first evaluation, i.e., we calculate cutoffs and accuracy of the new LITMUS tasks based on the identification of our clinical population by the use of norm-adjusted L1 and L2 tests. In particular, we want to know if the German SRT with the score of “target structure” can successfully identify bilingual children with SLI.

## Data Analysis

The children’s NWRT and SRT responses were recorded with special audio recorders. They were transcribed offline, verified and scored by two independent linguistically trained research and student assistants.

IBM SPSS 22 (2013) was used for all statistical analyses. Due to unequal group sizes and since explorative statistics revealed a violation of the assumption of normality in our data set, non-parametric statistical tests were used for group comparisons throughout the study. To measure the diagnostic accuracy of the LITMUS NWRT and SRT, sensitivity (the proportion of children with SLI identified as such by the task) and specificity (proportion of children with typical language development identified as such by the task) were calculated for each task upon an established cutoff score. The optimal cutoff score on a test is the performance score yielding the highest specificity and sensitivity ratios. Sensitivity and/or specificity rates  $\geq 90\%$  are considered good, whereas rates between 80 and 89% are considered fair (Plante and Vance, 1994). In addition, likelihood ratios were calculated for the established sensitivity and specificity levels because they are less likely to be affected by variations in the sample’s characteristics (see Dollaghan, 2004). A positive likelihood ratio (LR+) indicates the likelihood of scores below a cutoff criterion to occur in children with language impairment and is calculated as follows:  $LR+ = \text{sensitivity} / (1 - \text{specificity})$ . The negative likelihood ratio (LR−), on the other hand, indicates

the likelihood of a child performing above the cutoff point to be typically developing and is calculated with the following formula:  $LR- = (1 - \text{sensitivity}) / \text{specificity}$ . LR+ values  $\geq 10$  are considered to be clinically informative (highly indicative) of the presence of an impairment, and LR− values  $\leq 0.10$  are viewed as highly indicative of the absence of impairment.  $LRs+ \geq 3.0$  and  $LRs- \leq 0.3$  are viewed as “clinically suggestive,” whereas  $LRs+ < 3.0$  and  $LRs- > 0.3$  are considered to be clinically uninformative (e.g., Dollaghan, 2007).

Receiver operating characteristic (ROC)<sup>10</sup> curve analysis (Dunn, 2011) is widely used to estimate the discriminatory power and optimal cutoff criterion of a task. The optimal cutoff point is the score associated with the highest diagnostic accuracy of a task and is generated by plotting “the true positive rate (sensitivity) against the false positive rate (1 − specificity)” (Gutiérrez-Clellen and Simon-Cereidjido, 2010). One of the important drawbacks of the ROC analysis is that it uses the dichotomous variable “clinical group membership” as dependent variable to predict sensitivity and specificity for different thresholds. Thus, sensitivity and specificity ratios obtained by this procedure could be influenced by how well the participants were assigned to the SLI and TD groups. Since the clinical status of the bilingual children was determined using norm-referenced L1 and L2 tests standardized on monolingual children with adapted bilingual cutoffs, one cannot fully rule out the possibility of false group assignment especially in cases of selective impairments. For the aforementioned reasons, ROC curve analysis was performed only for our monolingual data. In case of bilinguals, we opted for an alternative measure that does not rely on the assignment procedure. We use *k*-means cluster analysis, which is one of the simplest clustering algorithms, to partition data into *k* clusters (MacQueen, 1967). The *k*-means clustering algorithm attempts to show which cluster each observation belongs to. In our case, the algorithm classified our observations into two clusters using the test variables as dependent measures. Crucially, such clusters are extracted based on the mathematical characteristics of the data independently from clinical status in an unsupervised manner, that is, assigned clinical status is not taken into consideration in the clustering procedure.<sup>11</sup> We ran *k*-means cluster analyses on each of the LITMUS tasks separately entering just one dependent measure into the clustering procedure at each run.

Our premise was that the two clusters would cut across the clinical status, since our test variables (LITMUS NWRT and SRT) have been proposed to be sensitive to the presence or absence of language impairment, see Section “The Language Impairment Testing in Multilingual Settings (LITMUS) Tools for Crosslinguistic Research.” The cutoff is a reference value ascertained after the cluster memberships are determined. Since we have uni-dimensional data (using just one variable per cluster analysis), the cutoff is on the same scale as the score of the dependent measure. The cutoff is an imaginary line separating the two clusters. It is calculated as the mean of the maximum score in

<sup>10</sup> A ROC analysis is currently being prepared for the performance of the bilingual groups on NWRT and in SRT-Id.

<sup>11</sup> We thank Istvan Fekete for drawing our attention to this method and his support with statistics in the following analysis.

**TABLE 6** | Summary of bilingualism factors in the bilingual groups [mean (SD) and range].<sup>a</sup>

|                                        | BiTD                       |                            |                             |                            | BiSLI ( <i>n</i> = 8)      |
|----------------------------------------|----------------------------|----------------------------|-----------------------------|----------------------------|----------------------------|
|                                        | BiTD-A ( <i>n</i> = 8)     | BiTD-P ( <i>n</i> = 19)    | BiTD-T ( <i>n</i> = 19)     | Total ( <i>n</i> = 46)     |                            |
| Age of onset                           | 29.88 (13.18)<br>0–42      | 15.74 (17.46)<br>0–39      | 22.58 (16.25)<br>0–48       | 21.02 (16.77)<br>0–48      | 19.50 (16.59)<br>0–36      |
| Length of exposure                     | 58.62 (19.17)<br>36–99     | 67.10 (22.48)<br>34–101    | 66.10 (21.48)<br>34–111     | 65.21 (21.30)<br>34–111    | 62.00 (22.77)<br>30–88     |
| Early L1 exposure (%)                  | 83% (19.9)                 | 70% (24.3)                 | 67% (19.6)                  | 71% (22.0)                 | 71% (11.97)                |
| Early L2 exposure (%)                  | 36% (30.9)                 | 57% (21.2)                 | 45% (27.0)                  | 48% (26.1)                 | 71% (10.68)                |
| Current L1 richness (/14)              | 3.5 (1.41)                 | 4.0 (1.5)                  | 4.47 (2.41)                 | 4.11 (1.94)                | 4.0 (0.0)                  |
| Current L2 richness (/14)              | 9.63 (1.76)                | 9.32 (1.76)                | 10.16 (2.41)                | 9.72 (2.05)                | 9.0 (0.0)                  |
| Language Dominance Index (–50 ... +50) | –3.63 (11.2)<br>–26 to +11 | 1.63 (12.66)<br>–17 to +23 | –4.97 (12.53)<br>–28 to +18 | –2.01 (12.5)<br>–28 to +23 | 1.13 (13.15)<br>–21 to +24 |

<sup>a</sup>When applicable.

the “lower” cluster and the minimum score of the “higher” cluster. Individual data points (here scores) allotted to the participants can then be ordered by group, which in turn allows calculation of sensitivity and specificity of the test.

## RESULTS

### Background Comparisons on Bilingualism Measures

In Section “Participants and Procedure for Verification of Clinical Status,” we established that the bilingual groups were comparable in terms of the LI variables “age, non-verbal intelligence and SES.” We further compared the bilingual groups for language background information obtained *via* the PaBiQ as displayed in **Table 6**. Group comparisons using a non-parametric Kruskal–Wallis test revealed no significant differences between the bilingual typically developing children according to L1 group on AoO, LoE, early L1 exposure, early L2 exposure, current L1 richness, and current L2 richness as well as the degree of L2 dominance as indicated by the LDI. Likewise, there were no significant differences between the BiSLI and BiTD groups on the aforementioned bilingualism measures.

Following the procedure and using the calculations described in Section “The LITMUS–PaBiQ,” we established language dominance in our groups of bilingual participants. **Table 7** summarizes these classifications by L1 and by final status. Note that in the Turkish/German typical children we find the highest rate of L1-dominant children. Among the BiSLI children, balanced or German dominant children are the majority. This might be a reflex of the traditional advice given to parents of bilingual children with language difficulties that they should use the majority language at home or with the child.

### Overall Results on the LITMUS NWRT and SRT

We first ran omnibus Kruskal–Wallis tests using scores on NWRT, SRT “identical repetition,” henceforth SRT\_Id, and SRT “target structure,” henceforth SRT\_Tar, as dependent variables to determine if clinical group has an effect. All three tests yielded significant results [ $\chi^2(3, N = 76) = 33.394, p < 0.001$  for NWRT,  $\chi^2(3, N = 76) = 38.926, p < 0.001$  for SRT\_Id, and

**TABLE 7** | Language dominance in bilingual children per L1.

|            | Final status | Dominant L1 | Balanced | Dominant L2 | Total |
|------------|--------------|-------------|----------|-------------|-------|
| Arabic     | BiTD         | 3           | 3        | 2           | 8     |
|            | BiSLI        | 0           | 2        | 1           | 3     |
| Portuguese | BiTD         | 8           | 3        | 8           | 19    |
|            | BiSLI        | 0           | 1        | 0           | 1     |
| Turkish    | BiTD         | 9           | 5        | 5           | 19    |
|            | BiSLI        | 2           | 1        | 1           | 4     |
| Total      | BiTD         | 20          | 11       | 15          | 46    |
|            | BiSLI        | 2           | 4        | 2           | 8     |

$\chi^2(3, N = 76) = 38.126, p < 0.001$  for SRT\_Tar]. In a next step, *post hoc* Mann–Whitney *U* comparisons were carried out on the dependent measures applying Bonferroni-adjustment of *p*-values to reduce Type I error that can arise due to multiple comparisons.

The overall performance of the different groups defined in **Table 2** in the NWRT and SRT is given in **Figure 1**. The NWRT significantly distinguishes the MoSLIs from the MoTDs ( $U = 5.5, p < 0.001, r = 0.767$ ) and the BiSLIs from the BiTDs ( $U = 24.5, p < 0.001, r = 0.528$ ). Moreover, BiTDs perform significantly different from the MoSLIs in the NWRT ( $U = 38.0, p < 0.001, r = 0.600$ ). This means that the LITMUS NWRT can identify SLI across populations. In addition, performance in NWRT does not statistically differ in BiTDs and MoTDs.

**Figure 1** further shows that the SRT can well discriminate SLI from TD children in monolingual and bilingual populations with both scoring methods. The score of SRT\_Id distinguishes the MoSLIs from the MoTDs ( $U = 0.000, p < 0.001, r = 0.846$ ) and the BiSLIs from the BiTDs ( $U = 36.00, p < 0.001, r = 0.490$ ). Here as well, BiTDs perform significantly better than the MoSLIs ( $U = 46.5, p < 0.001, r = 0.578$ ). If SRT is rated with the measure of SRT\_Tar, bilingual children perform better. Again, MoTDs are significantly different from MoSLIs ( $U = 0.000, p < 0.001, r = 0.844$ ), BiTDs perform significantly better than BiSLIs ( $U = 32.5, p < 0.001, r = 0.539$ ), and also BiTDs perform significantly better than MoSLIs ( $U = 40.0, p < 0.001, r = 0.595$ ). However, the MoTDs and BiTDs do not perform alike in the SRT by score SRT\_Id: ( $U = 76.5, p = 0.006, r = 0.438$ ) and SRT\_Tar: ( $U = 102.5, p = 0.036, r = 0.364$ ). Outliers in the SRT scored by

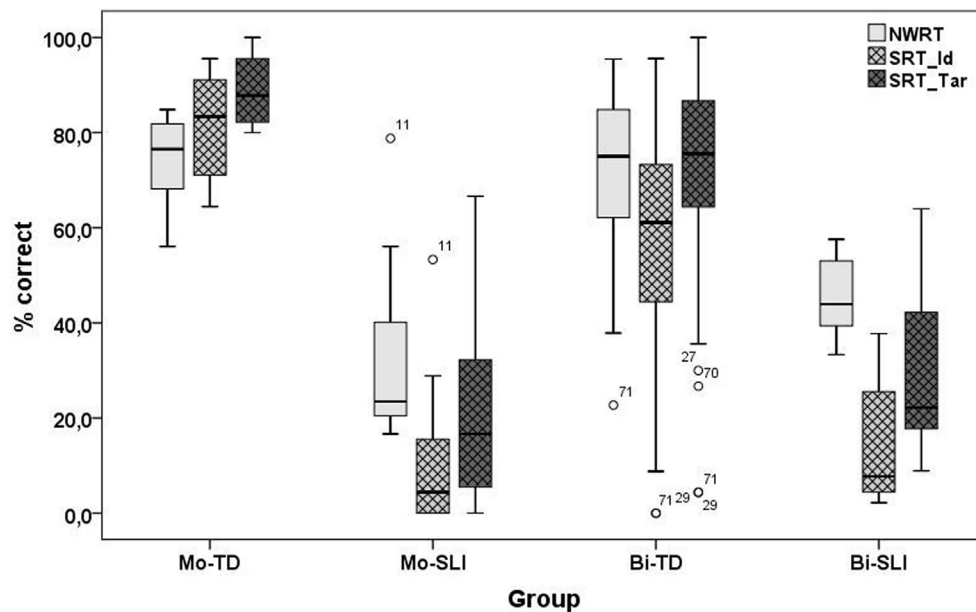


**FIGURE 1** | Non-word repetition task (NWRT), SRT\_Id, and SRT\_Tar: % of correct identical repetition split by group.

SRT\_Id are 29<sup>12</sup> and 71, where the latter is also the outlier in the NWRT. These two children perform within or even below the BiSLI range. The outlier in the MoSLI group, 11, is an older child (9;4). Bilingual children performing below the group range in the mastery of SRT\_Tar are 29 and 71, but also 27 and 70.<sup>13</sup>

For further analyses, we first present results from NWRT and SRT\_Id and single out SRT\_Tar for closer analysis. We first run a ROC curve analysis on MoTD and MoSLI to determine the optimal monolingual cutoff score and the diagnostic accuracy for each of the tasks. As can be seen in **Table 8**, both LITMUS tests have excellent diagnostic accuracy in monolingual children. When looking at the individual scores of the monolingual children, it emerges that a cutoff of 59.85% for the NWRT and 63.33% on SRT\_Id sharply group the children with 100% sensitivity and specificity for SRT\_Id and 91.7% sensitivity and 90% specificity for NWRT. Applying the measure SRT\_Tar to the monolingual data allows a cutoff of 77.78%, still with 100% sensitivity and specificity.

Comparison of the bilingual groups with the monolingual groups points to the fact that other factors than language impairment could lead to poor performance. To address this problem, we performed a *k*-means cluster analysis of the performance of all bilinguals on NWRT, SRT\_Id and SRT\_Tar. The *k*-clustering, unbiased as to any given classification of participants, renders two clusters, participants who are performing well on the task (cluster A “higher cluster”) and those performing poorly (cluster B “lower cluster”), the cutoff line between the two clusters was determined for each of the measures as outlined in Section “Data Analysis.”

<sup>12</sup>We use case numbers for the identification of individual participants.

<sup>13</sup>The outliers are included in the group analysis.

**TABLE 8** | Diagnostic accuracy of the German Language Impairment Testing in Multilingual Settings sentence repetition task (SRT) and non-word repetition task (NWRT) among monolingual children.

|         | AUC   | Cutoff (%) | Sensitivity   | Specificity  | LR+                    | LR–  |
|---------|-------|------------|---------------|--------------|------------------------|------|
| NWRT    | 0.954 | 59.85      | 91.7% (11/12) | 90% (9/10)   | 9.17                   | 0.09 |
| SRT_Id  | 1.000 | 63.33      | 100% (12/12)  | 100% (10/10) | Undefined <sup>a</sup> | 0.0  |
| SRT_Tar | 0.996 | 77.78      | 100% (12/12)  | 100% (10/10) | Undefined              | 0.0  |

<sup>a</sup>If specificity = 100% then LR+ ratios are undefined.

For the NWRT, the *k*-means cluster analysis rendered two clusters separated by a *k*-means cutoff of 63.5%: 34 children performing above cutoff and 20 children scoring below. In the SRT\_Id, the analysis rendered a 41.25% cutoff separating the two clusters. On this measure, 35 children, cluster A, scored above the cutoff, whereas 19 children, cluster B, performed below cutoff score. To complete the analysis and calculate the sensitivity and specificity levels, the individual values in the clusters for each task were identified as scores of individual BiTD or BiSLI children. **Figures 2** and **3** depict the performance of cluster A and cluster B in NWRT and SRT\_Id, respectively. All of the eight children assigned to the BiSLI group based on standardized test procedures belonged to the lower cluster on both measures, which yields a sensitivity of 100% with an LR– of 0.0. However, the specificity levels and the corresponding LR+ values for ruling out language impairment were only suggestive as can be seen in **Table 9**. This is ascribed to the fact that 12 children with the final status BiTD scored below cutoff on NWRT and 11 children scored below cutoff on SRT\_Id, and thus belonged to cluster B.

For the measure of SRT\_Tar, *k*-clustering resulted in two clusters separated by a 52.2% cutoff: 39 children, cluster A, performing above cutoff and 15 children, cluster B, performing

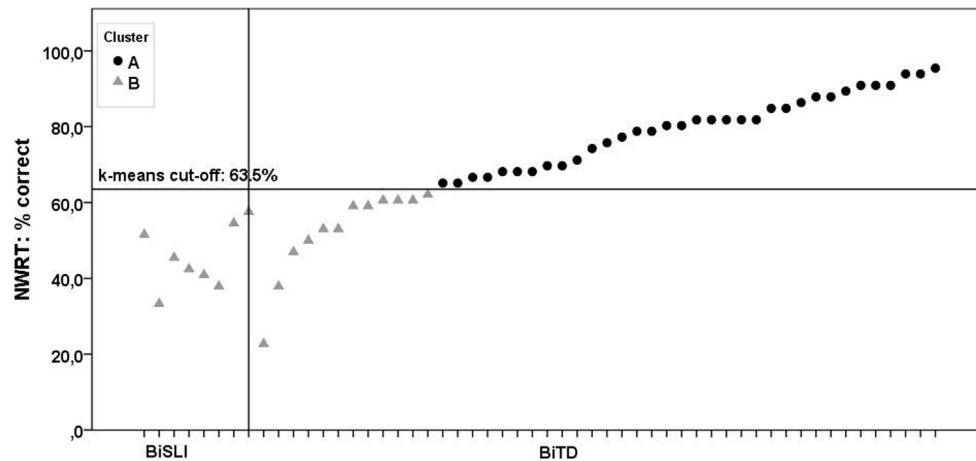


**FIGURE 2** | k-Means cluster analysis of performance of bilingual children on non-word repetition task (NWRT).

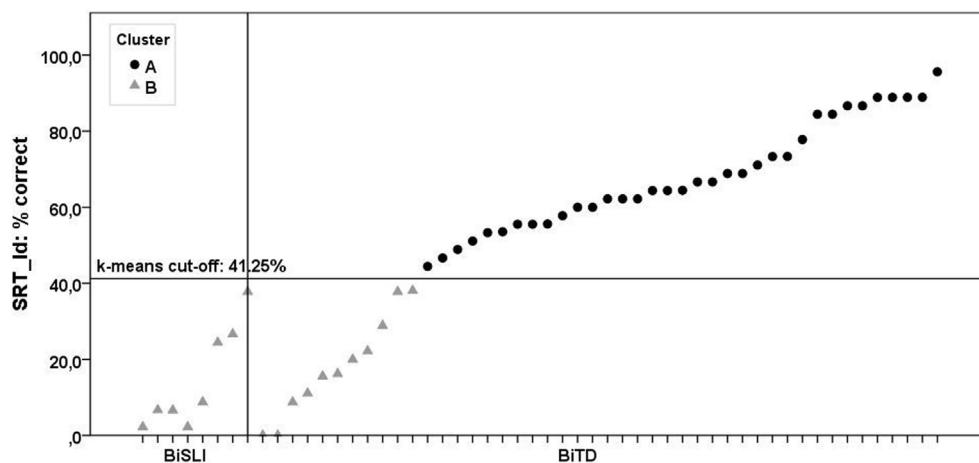


**FIGURE 3** | k-Means cluster analysis of performance of bilingual children on SRT\_Id.

**TABLE 9** | Diagnostic accuracy of the German Language Impairment Testing in Multilingual Settings sentence repetition task (SRT) and non-word repetition task (NWRT) among bilingual children for individual measures and for test combinations.

|                | Cutoff (%) | Sensitivity | Specificity    | LR+  | LR–  |
|----------------|------------|-------------|----------------|------|------|
| NWRT           | 63.5       | 100% (8/8)  | 73.91% (34/46) | 3.83 | 0.0  |
| SRT_Id         | 41.25      | 100% (8/8)  | 76.1% (35/46)  | 4.18 | 0.0  |
| SRT_Tar        | 52.2       | 87.5% (7/8) | 82.6% (38/46)  | 5.03 | 0.15 |
| NWRT + SRT_Id  |            | 100% (8/8)  | 86.95% (40/46) | 7.66 | 0.0  |
| NWRT + SRT_Tar |            | 87.5% (7/8) | 86.95% (40/46) | 6.70 | 0.14 |

below this cutoff. **Figure 4** shows the individual performance of members of cluster A and cluster B in the above measure. The children classified as BiSLI all belonged to cluster B, except **26** who is 9;1 years old and does not seem to be impaired in German morphosyntax. However, eight children in cluster B, some with extremely low scores, had received the final status of BiTD. These children are in particular: **70, 71, 44, 45, 27, 76, 28,**

and **29**. Interestingly, most of these children except for **44** and **45** performed below *k*-means cutoff on both SRT\_Id and NWRT. It remains to be investigated why these children scored low.

The measure SRT\_Tar, which gives more weight to mastery of syntactically complex structures than to lexical abilities, gave lower sensitivity but better specificity levels than SRT\_Id or NWRT, see **Table 9**. We also investigated whether combining the NWRT with SRT raises diagnostic accuracy. The results in **Table 9** indicate that a combination of NWRT and SRT\_Id or NWRT and SRT\_Tar indeed results in better specificity and thus overall diagnostic accuracy.

## Dominance As a Factor for the Performance of Bilingual Typically Developing Children

To examine whether language dominance affects the performance of bilingual children without language impairment, we plotted the children's individual scores on SRT\_Tar and NWRT against

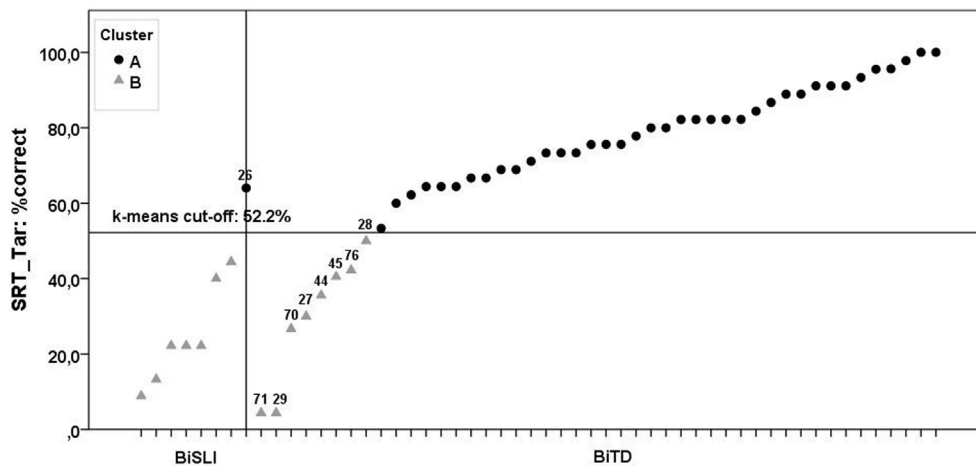


**FIGURE 4** | *k*-Means cluster analysis of performance of bilingual children on SRT\_Tar.

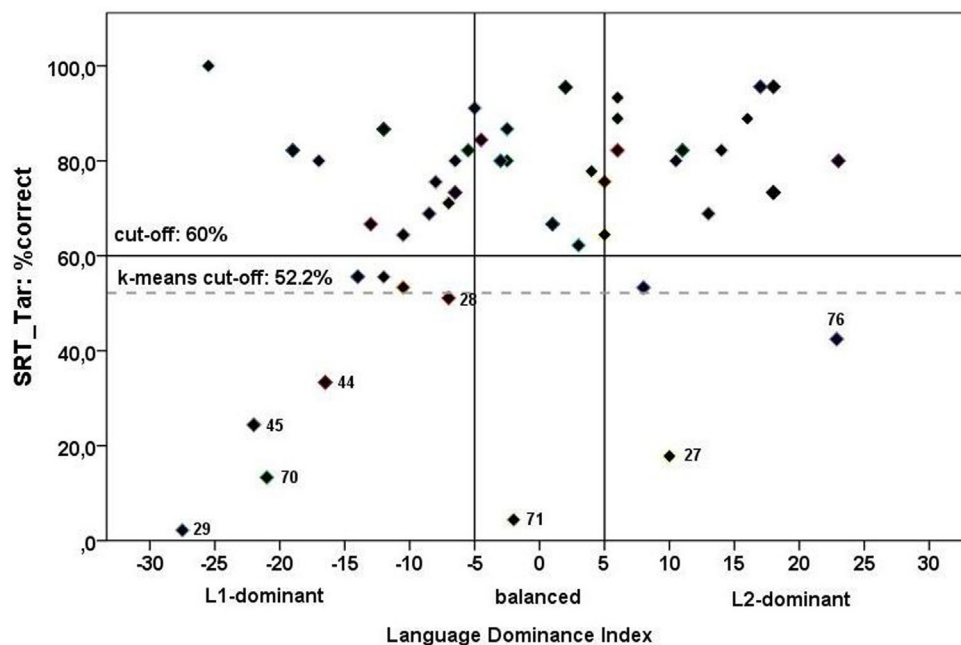


**FIGURE 5** | SRT\_Tar vs. Language Dominance Index: individual results of children classified as BiTD by L1 and L2 tests.

their LDI.<sup>14</sup> As illustrated in **Figure 5**, it emerges that language dominance strongly influences performance of the typical bilingual children in SRT\_Tar. On the other hand, just as Almeida et al. (2017) show for the French SRT, among the 20 L1-dominant children the majority, here 70% (14/20) score over 60% correct in SRT\_Tar and 75% (15/20) score over 52.2% (see **Figure 5**). The five L1-dominant children who perform below a 52.2% cutoff are identified as 70, 44, 45, 28, and 29.

At first glance (see **Figure 6**), language dominance may seem to influence performance on NWRT to the same extent as in

SRT\_Tar: 6 out of 20 L1-dominant children perform below the *k*-means cutoff score of 63.5%. However, unlike in the SRT\_Tar, four of the latter six children perform almost at cutoff ( $\geq 61\%$  correct) and all children perform above cutoff on the LD part of the task.<sup>15</sup> 29, who performs below cutoff on NWRT, scores above cutoff in the LD part. Only two L1-dominant children 50 and 28 had an overall score on NWRT  $< 61\%$  due to poor performance on the LD part of the NWRT (50: 27.78% correct, 28: 44%

<sup>14</sup>We chose only the measure with higher specificity for the SRT.

<sup>15</sup>In this study, the results of both LI and LD parts of the NWRT are collapsed together. However, in cases of L1-dominant children, we verified that their scores on the LD part were above cutoff to exclude the potential effect of L1 dominance.

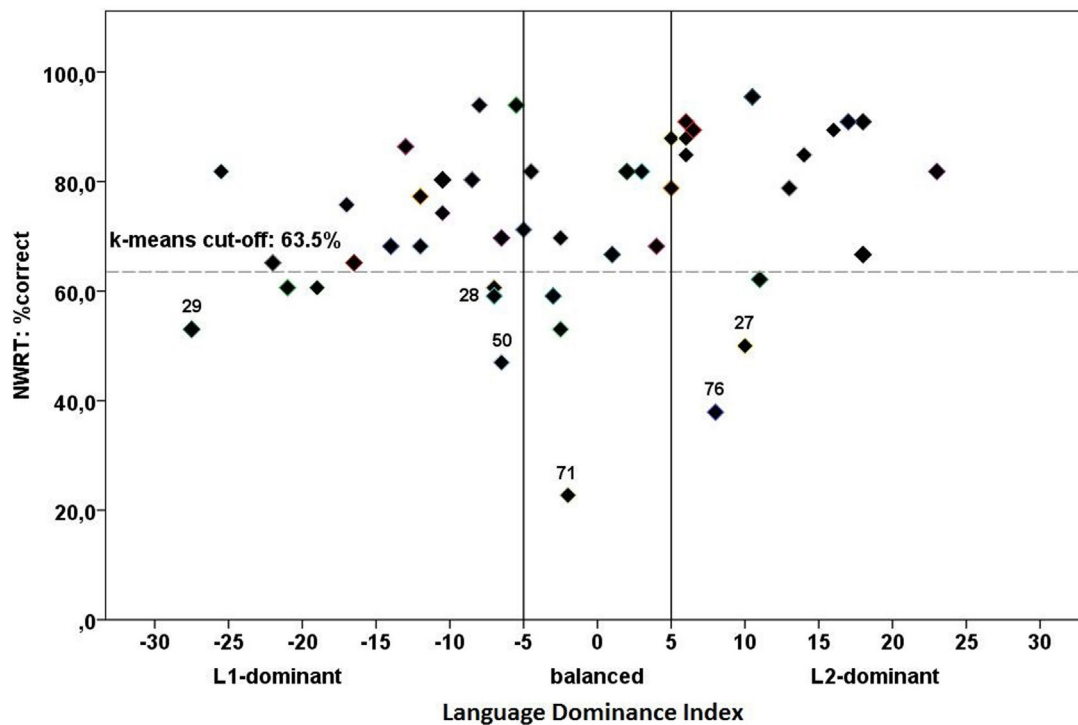


**FIGURE 6** | Non-word repetition task (NWRT) vs. Language Dominance Index: individual results of children classified as BiTD by L1 and L2 tests.

correct). This allows the conclusion that performance on the NWRT is less independent of language dominance than the SRT. Note also that among the balanced and German dominant children, only three score below the cutoff both in SRT\_Tar and in NWRT and these children are 71, 76, and 27, whose status might have to be reanalyzed as will be discussed in Section “Discussion.”

## DISCUSSION

### Summary

This study investigated the accuracy of two German LITMUS tasks, an NWRT and an SRT in the identification of language impairment in bilingual children. Both NWRT and SRT prove to have good sensitivity and specificity in monolinguals: NWRT (sensitivity = 91.7%, specificity = 90%) and SRT show 100% sensitivity and specificity for both scoring methods SRT\_Id and SRT\_Tar. The results for monolinguals clearly show that the tasks are well constructed and reliably identify SLI. The same can be said for bilingual settings. Especially, the fact that the results for the NWRT are more or less independent of language dominance makes it a valuable new tool for language assessment. The reduced specificity of the SRT for bilinguals is due to several factors that will emerge more clearly in a detailed discussion of the individual cases we highlighted in Sections “Overall Results on the LITMUS NWRT and SRT” and “Dominance As a Factor for the Performance of Bilingual Typically Developing Children.” It was noteworthy that the same children were identified in several types of analyses as either being “underdiagnosed” by the SRT (26) or of being

“overdiagnosed” (71, 27, and 76) by both SRT and NWRT. Moreover, L1-dominant children such as 70, 44, 45, 28, and 29 also performed under cutoff 52.2% in the SRT\_Tar.

In Section “Dominance As a Factor for the Performance of Bilingual Typically Developing Children,” we already identified one factor of possible misdiagnosis in L2 tasks, namely, L1 dominance, see Figure 5. L1 dominance will then interact with other factors, which we discuss in the following. One possibility for a reduced diagnostic accuracy is that our final status assignment might have been too strict, see also Bossuyt et al. (2015) on the impact of clinical group definition on accuracy measures. Note that Armon-Lotem and Meir (2016) use L1 and L2 tests with global scores, but additionally rely on parental or teacher concern. Boerma et al. (2015) and Boerma and Blom (2017) rely on clinical referral, i.e., on L2 testing exclusively. Thordardottir (2015) recommends including measures from samples of spontaneous production in addition to norm-referenced L1 and L2 tests. Given these different methods for identifying the clinical population, we will discuss cases of possible misclassification by our procedure, drawing also on impressions from the samples of narratives we have at our disposal. Alternatively, and given that the SRT, and SRT\_Tar in particular, targets morphosyntactic skills, misclassification could arise because our procedure did not take into account selective impairments such as grammatical/syntactic SLI. This would mean that an individual child has been classified as BiTD, but is syntactically impaired, which arguably leads to poor performance in SRT\_Tar. Children who show impairments in phonology and lexicon, but not in morphosyntax, would have been classified as BiSLI, but will not

necessarily perform poorly in the SRT. Finally, misclassification could arise if standardized tests are not reliable in certain constellations of bilingualism, such as heritage situations.<sup>16</sup> In the discussion, we specifically address the problems arising from our strict procedure and the (non)-applicability of standardized L1 tests in heritage situations.

## Subgroups of SLI

Since our classifying procedure did not isolate subgroups of SLI, but clearly aimed at a broader definition, we first address this problem by discussing the cases revealed by the clustering for SRT\_Tar, see **Figure 4**. The BiSLI child in cluster A, **26**, was classified as BiSLI because of her scores in the L1 test, ELO-L, and because her lexical and phonological abilities were below norm in L2. Note that she was 9;1 years at the time of testing but she performed well below the norms for younger children (7;11) in the L1 test, in which her sentence production showed a slight impairment whereas her phonological production showed great deficits. For L2 testing, the lexical test is normed till 9;11, for phonology she was below norm of younger children and the LiSe-DaZ norms could be age adjusted as described in Section “Standardized L2 and L1 Tests.” Her spontaneous L2-language sample did not evidence any of the characteristic markers of SLI. This indicates that she might not be syntactically impaired. This seems to be confirmed by her good score in SRT\_Tar. Not surprisingly, her score in SRT\_Id was below 41.25%, despite her age, since it involves recollection of vocabulary.

The same problem, namely, that the impairment might be selective, is exemplified by **27**, who is BiTD because of reasonable scores in L2 lexicon and phonology, but has clear problems in some areas of morphosyntax identified by the LiSe-DaZ, among them SVA. Since this domain does not receive a *t* value in the test, it is not included in our final evaluation procedure. **27** is a simultaneous bilingual, L2 dominant, clearly impaired in L1, and her spontaneous production in both languages confirms problems with morphosyntax. **27** performs low in the SRT in both measures as well as in the NWRT and hence may be selectively impaired, i.e., the final status allotted may be misleading.

To see whether we might have missed bilingual children with grammatical SLI, we first consider children who show poor L2 performance only in the LiSe-DaZ. These are **24**, **27**, and **28**. **27** was already discussed as a possible case of grammatical SLI. **28** is below cutoff in SRT\_Tar and in the NWRT. She is also language impaired by her L1 status and her parents voiced concern. In other words she would be a BiSLI child if we had included selective impairments. **24** does not show problems in any of the experimental tasks and is not L1 impaired.

To summarize: **27** and **28** might be cases of grammatical SLI who show poor performance in LiSe-DaZ, but also in the SRT. On the other hand, **26** might be a case of lexical/phonological impairment with good performance in the LiSe-DaZ and in SRT\_Tar.

Reevaluating L1-dominant children below cutoff (52.2%) by taking into account the possibility of selective impairments, leads to the following picture: **70**, **45**, and **29**<sup>17</sup> have a very high L1-language index and score as typical children in their L1 tests. **45** also performs well in the NWRT, and **70** performs almost at cutoff on the task. This implies that L1 dominance explains the performance of **70** and **45**. **28** and **44**, however, score as language impaired in their L1 and are classified as impaired in one of the L2 tests applied here: **28** has morphosyntactic problems, **44** performs low in the lexical assessment. These children might therefore be selectively impaired.<sup>18</sup> However, only **28** also performs low in the NWRT, whereas **44** performs above cutoff. **29** remains a problem since she performs low in all L2 tests as well as the experimental tasks whereas her L1 test puts her firmly among the typical children.

Interestingly, most of the children discussed above perform below cutoff not only in SRT, but also in NWRT: **71**, **76**, and **27** among the balanced and L2-dominant children, and **28** and **29** among the L1-dominant children.

## L1 Assessment in Heritage Situations

It is not surprising that some children who perform below cutoffs in the standardized L2 tasks (and also in the SRT and NWRT), nevertheless have a final status as BiTD because they did well in the L1 tests. Five of the 46 typical bilinguals would be diagnosed as language impaired by the L2 tests but are doing well in L1. Especially for L1-dominant children this might be expected. **29** is a case in point: all three German tests classified the child as impaired, so would NWRT and SRT. The child did perfectly well in the TEDIL, however. **71**, who is balanced according to the PaBiQ, seems to be a similar case but turns out to be a child whose final status might be reconsidered: **71** scored above the norms in lexical reception in the GOL-E, but would have been below the norm in the assessment of the lexicon provided by the PALPA-P, which exists for her age range. Recall that we decided to use GOL-E as lexical assessment for all Portuguese children because in this area there were age gaps in the norms for the PALPA-P, which does not apply to **71**. More surprising is the fact that even German dominant children who scored as impaired in the German tests sometimes do well in the L1 tasks as is the case of **76**. However, **76** performed only minimally above cutoff in the TEDIL. Both **76** and **71** would be classified as language impaired if their (L1 and L2) samples of spontaneous production had been included in the initial decision about final status.

If the results of the L1 tests are examined more closely, it is rather striking that 16 of the 46 bilingual typical children have an L1 diagnosis of impairment, whereas only 5 are so diagnosed by the L2 tests. Examining these numbers by dominance we see that among the 20 L1-dominant children, 6 would have been diagnosed as impaired by the L1 test (1 Arabic child, 3 Portuguese children, and 2 Turkish children). Among the 11 balanced

<sup>16</sup>Our full test battery included WM tasks and tasks measuring executive function. We measured forward digit span (FDS) and backwards digit span as WM measures. Preliminary regression analyses showed that FDS only explains a small portion of the variance in SR performance in typical bilinguals. Therefore, the possible influence of WM on the performance in these tasks is not further pursued here.

<sup>17</sup>With a 60% cutoff (**Figure 5**), **47** would also be below cutoff. This child is L1 dominant and performs below norms in L2-lexical skills but within norms in the L1 assessments and in the NWRT.

<sup>18</sup>**73** and **26** were classified as BiSLI by the strict criteria, and like **44**, do not show a morphosyntactic impairment, but are impaired in the lexicon in particular.

children, 2 would have been diagnosed as L1 impaired. Among the 15 German dominant children, 8 would have been diagnosed as L1 impaired (2 Arabic children, 5 Portuguese children, and 1 Turkish child). These figures point to problems with the applicability of the L1 tests, which, in turn, call in question the final BiSLI status.

There are multiple reasons for this situation. Heritage speakers growing up as simultaneous bilinguals often show differences to monolingual speakers in their adult performance. This seems to concern morphosyntax and lexicon more than phonology (Montrul, 2010; Rinke and Flores, 2014). Reasons for this situation have been sought in the fact that children who have been exposed to their L2 early or are simultaneous bilinguals are often subject to language attrition in their L1 or could be claimed to suffer from incomplete acquisition (Köpke et al., 2004; Montrul, 2008, 2010; Benmamoun et al., 2013). Moreover, the language of children growing up as Turkish/German bilinguals in Germany is special from several perspectives: They are often third generation heritage speakers and Immigrant Turkish in Germany has features (Schroeder and Dollnick, 2013) which count as clinical markers for SLI in Standard Turkish (see also Chilla and San, 2017). Finally, the L1 tests we chose might have other inherent problems: The TEDIL only has two global scores, which do not allow identifying language domains as specifically problematic. The version of the Portuguese PALPA-P that we used has been normed with only few children for some ages and subtests. It is a linguistically well-controlled test but the lack of norms in receptive vocabulary in crucial age ranges made it necessary to use a different test for assessment of the lexicon, the GOL-E. Some of the misdiagnosis may therefore be due to the specific language tests chosen here. The more fundamental problems seem to be the heritage situation and language attrition of L1 which has been shown to be particularly noticeable when L2 exposure is early, see Lein et al. (2017) for an analysis of heritage effects in the Portuguese bilinguals also investigated in this study.

If the 15 children dominant in their L2 German are considered, more than half of them (8) would have been classified as SLI if only tested in their L1, which is not surprising. In contrast, overdiagnosis due to the L2 tests did not occur. German dominant children with a final classification of BiTD (also considering L1) were all correctly classified as BiTD by the combination of the three norm-referenced tests used for classification. Of the 46 children with a final classification of BiTD there are only three children who would have been BiSLI if only the L2 had been considered. This seems to indicate that in heritage situations L2 tests are more reliable than L1 tests, which may have multiple reasons: the contact situation and the existence of immigrant varieties, language attrition, and possibly properties of the L1 tests. Incidentally, 29, who remains a problematic case also after we reconsidered the status of the bilingual children, might highlight the problems with the Turkish test, see also Almeida et al. (2017).

## Reconsidering the Status of the Bilingual Children

Following the argumentation about (a) selective impairments and (b) possible problems with L1 tests in heritage situations,

we suggest different criteria for the identification of the bilingual clinical group: We consider children as BiSLI if they have a selective L2 impairment, and score below norms in their L1 tests or show poor spontaneous production in both languages.<sup>19</sup> These criteria still require an impairment in both languages of a bilingual child but would classify 71, 76, 27, 28, and 44 as BiSLI. Incidentally, two of these children had been in SLT when recruited (27 and 28) and the remaining three might be cases of underdiagnosis. Given the clustering shown by Figures 2–4, and the foregoing discussion of these particular cases, such a regrouping would clearly raise diagnostic accuracy for all measures (SRT\_Id, SRT\_Tar and NWRT).

Given that the grouping of children we presented in Table 2 takes into account language dominance by adjusting the norms in standardized L1 and L2 tests, the discussion points raised above show that especially selective impairments should be taken into account when deciding on the status of a bilingual child and when considering the accuracy of a particular test, which might be targeting one language domain more than others. The heritage situation adds to the difficulty and the cases discussed suggest that norm adjustments for L1 might have to be reconsidered for heritage speakers.

## CONCLUSION

Our investigation of the German LITMUS NWRT and SRT has shown that both are well suited as tools for the identification of SLI in bilinguals. Both tasks clearly identify SLI in German monolinguals demonstrating that they target crucial phonological and syntactic areas and structures. In addition, both tasks can identify SLI in bilingual contexts. Since the construction of both tasks was guided by linguistic notions such as phonological or syntactic complexity and neither task primarily measures WM, this is a result relevant on the theoretical and the practical level.

Both tasks clearly measure linguistic abilities, the NWRT on the phonological side, the SRT in morphosyntax. The SRT was scored in two different ways: SRT\_Id as a measure includes all morphosyntactic but also all lexical errors, not cumulating them. SRT\_Tar scores only morphosyntax and, by concentrating on syntactic structures and not counting morphological errors such as case or gender if they do not change the structure aimed at, does not penalize bilingual children and seems a good measure of (morpho)syntactic abilities. From the practical point of view, the possibility of using both or one of these scoring methods allows fine-grained diagnosis of the impaired domains. Concentrating on certain structures such as those involving Wh-movement with and without embeddings or intervention (see The German LITMUS SRT) would give an even more detailed picture but was not the focus of this study.

Our evaluation of the LITMUS tools started with rather strict criteria as to the status of a bilingual child as typical or language impaired. We classified a child as BiSLI only if the

<sup>19</sup>We do not apply any formal measure here but judge production by certain markers: correct SVA, sentence bracket or V2 and presence of embeddings.

child scored below (adjusted) norms in two domains of both her L1 and her L2. For this categorization, and also for further evaluation of our results, see Section “Dominance As a Factor for the Performance of Bilingual Typically Developing Children,” the parental questionnaire, the PaBiQ, was an indispensable tool. We concentrated on the language dominance value, which allowed adjusting the norms for standardized tests and helped us in the interpretation of our results on the performance in the LITMUS NWRT and SRT. It emerged that performance in the NWRT is largely independent of language dominance, whereas it influences performance in the SRT. However, 75% of the L1-dominant children performed above the cutoff in the SRT when scored as SRT\_Tar, so that accuracy remains satisfactory. Similar findings are reported in Almeida et al. (2017) for the corresponding French tasks and in Grimm and Hübner (in press) for the German NWRT. Interestingly, LoE does not influence performance in the NWRT either, as reported in Grimm and Hübner (in press).<sup>20</sup>

Considering individual cases and their performance in these new tasks revealed that the grouping we chose on the basis of standardized L1 and L2 tests might have missed cases of language impairment, which would not be surprising giving that this classification cut in half the group of children in SLT. We attributed such missed cases to either the problems of using and interpreting L1 tests (even with adjusted norms) in heritage contexts or to cases of selective language impairments. Clearly, interpretation of individual results for bilinguals is impossible without background information as provided by parental questionnaires.

On the practical level, this leads us to conclude that the LITMUS NWRT and SRT are indeed reliable tools that can be used as a first evaluation of a child's language abilities, singly, but better in combination. Since their administration takes only a fraction of the time that has to be invested for standardized tests, this is a good overall result. On the theoretical level, we have shown that L2 tasks, if linguistically well controlled and targeting complex structures, clearly identify language impairment in bilingual contexts.

<sup>20</sup>Since this factor contributes to our dominance calculation, we did not consider it separately.

## REFERENCES

- Abed Ibrahim, L., and Hamann, C. (2017). “Bilingual Arabic-German & Turkish-German children with and without specific language impairment: comparing performance in sentence and nonword repetition tasks,” in *Proceedings of BUCLD 41*, eds M. LaMendola and J. Scott (Somerville: Cascadia Press), 1–17.
- Abed Ibrahim, L., Lein, T., Rothweiler, M., and Hamann, C. (in press). “SLI in bilinguals: comparing LITMUS-tasks to a standardized test with bilingual norms,” in *Bilingualism and Specific Language Impairment, Bi-SLI 2015*, eds C. dos Santos and L. de Almeida (Amsterdam: Benjamins).
- Almeida, L., Ferré, S., Morin, E., Prévost, P., dos Santos, C., Tuller, L., et al. (2017). Identification of bilingual children with specific language impairment in France. *Linguist. Approaches Biling.* 7, 3/4. doi:10.1075/lab.15019.alm
- Archibald, L., and Gathercole, S. (2006). Short-term and working memory in specific language impairment. *Int. J. Lang. Commun. Disord.* 41, 675–693. doi:10.1080/13682820500442602

## ETHICS STATEMENT

This study was carried out with the recommendation of the “Kommission für Forschungsfolgenabschätzung und Ethik” (Commission for the Evaluation of Research Consequences and Ethics) of the Carl-von-Ossietzky University of Oldenburg (ref. Drs. 21/16/2013) with written informed consent from all subjects. All subjects (or their parents) were informed that audio recordings were made and gave written consent in accordance with the Declaration of Helsinki. The protocol was approved by the Kommission für Forschungsfolgenabschätzung und Ethik of the Carl-von-Ossietzky University of Oldenburg.

## AUTHOR CONTRIBUTIONS

Both the authors CH and LA are fully responsible for all parts of the text and the analyses. LA collected most of the data and conducted its analysis.

## ACKNOWLEDGMENTS

This study was funded by DFG (German Science Foundation) grant HA 2335/6-1 to Cornelia Hamann. It is part of the BiLaD project with additional DFG grants to Monika Rothweiler and Solveig Chilla as well as an ANR (French Science Agency) grant to Laurice Tuller and her team. The authors thank all the investigators involved in the project for their continued advice and support, with special thanks to Tatjana Lein and Hilal San for collecting and analyzing the Portuguese/German and the Turkish/German data. Special thanks also go to Angela Grimm for sharing her German LITMUS NWRT and to Istvan Fekete for theoretical and practical support with the statistical analysis. The authors thank all the parents, educators, teachers, and speech language therapists for their cooperation and, last but not least, the authors particularly thank the children for their participation and their patience with us in completing the tasks.

## FUNDING

This work was funded by DFG (German Science Foundation) grant HA 2335/6-1 to CH, supplying funding for the PhD work of LA.

- Armon-Lotem, S., and Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *Int. J. Lang. Commun. Disord.* 51, 715–731. doi:10.1111/1460-6984.12242
- Armon-Lotem, S., de Jong, J., and Meir, N. (2015). *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters.
- Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: opportunities and challenges for linguistics. *Theor. Linguist.* 39, 129–181. doi:10.1515/tl-2013-0009
- Boerma, T., and Blom, E. (2017). Assessment of bilingual children: what if testing both languages is not possible? *J. Commun. Disord.* 66, 65–76. doi:10.1016/j.jcomdis.2017.04.001
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., and Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning dutch as a second language. *JSLHR* 58, 1747–1760. doi:10.1044/2015\_JSLHR-L-15-0058

- Bossuyt, P., Reitsma, J., Bruns, D., Gatsonis, C., Glasziou, P., Irwig, L., et al. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* 277, 826–832. doi:10.1148/radiol.2015151516
- Bulthell, S., and Häcker, H. (2002). *Coloured Progressive Matrices*. Frankfurt: Pearson Assessment.
- Castro, S. L., Caló, S., Gomes, I., Kay, J., Lesser, R., and Coltheart, M. (2007). *PALPA-P Provas de Avaliao da Linguagem e da Afasia em Portugues*. Lisbon: CEGOC.
- Chiat, S. (2015). “Non-word repetition,” in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 123–148.
- Chiat, S., and Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: effects of bilingualism and socioeconomic status on children's performance. *JSLHR* 59, 1179–1189. doi:10.1044/2016\_JSLHR-L-15-0293
- Chilla, S., and San, H. (2017). “Möglichkeiten und Grenzen der Diagnostik erstsprachlicher Fähigkeiten: Türkisch-deutsche und türkisch-französische Kinder im Vergleich,” in *Sprachen 2016: Russisch und Türkisch im Fokus*, eds C. Yildiz, N. Topaj, R. Thomas, and I. Gülzow (Berlin: Peter Lang Verlag), 175–205.
- Clahsen, H. (1991). *Child Language and Developmental Dysphasia: Linguistic Studies of the Acquisition of German*. Amsterdam: John Benjamins.
- Cobo-Lewis, A., Pearson, B., Eilers, R., and Umbel, V. (2002). “Effects of bilingualism and bilingual educations on oral and written English skills: a multifactor study of standardized test outcomes,” in *Language and Literacy in Bilingual Children*, eds D. K. Oller and R. E. Eilers (Clevedon: Multilingual Matters), 43–63.
- Conti-Ramsden, G., Botting, N., and Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *J. Child Psychol. Psychiatry* 42, 741–748. doi:10.1111/1469-7610.00770
- Cornips, L., and Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Lang. Res.* 24, 267–297. doi:10.1177/0267658308090182
- Dollaghan, C. (2004). Evidence-based practice in communication disorders: what do we know, and when do we know it? *J. Commun. Disord.* 37, 391–400. doi:10.1016/j.jcomdis.2004.04.002
- Dollaghan, C. (2007). *The Handbook for Evidence-Based Practice in Communication Disorders*. Baltimore, MD: Paul H. Brookes.
- dos Santos, C., and Ferré, S. (2016). A non-word repetition task to assess bilingual children's phonology. *Lang. Acquis.* 1–14. doi:10.1080/10489223.2016.1243692
- Dunn, G. (2011). “Statistics in psychiatry,” in *International Encyclopedia of Statistical Science*, ed. M. Lovric (Berlin: Springer), 1136–1138.
- Ferré, S., Tuller, L., Sizaret, E., and Barthez, M.-A. (2012). “Acquiring and avoiding phonological complexity in SLI vs. typical development of French: the case of consonant clusters,” in *Consonant Clusters and Structural Complexity*, eds P. Hoole, L. Bombine, M. Pouplier, C. Mooshammer, and B. Kühnert (Berlin: de Gruyter), 285–308.
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., and Zebib, R. (2016). How to identify SLI in bilingual children: a study on sentence repetition in French. *Lang. Acquis.* 1–17. doi:10.1080/10489223.2016.1192635
- Fox-Boyer, A. V. (2014). *Psycholinguistischer Analyse kindlicher Aussprachestörungen-II: PLAKSS II*. Frankfurt am Main: Pearson.
- Fredman, M. (2006). Recommendations for working with bilingual children – prepared by the multilingual affairs committee of IALP. *Folia Phoniatri. Logop.* 58, 458–464. doi:10.1159/000096570
- Friedmann, N., and Novogrodsky, R. (2011). Which questions are most difficult to understand? The comprehension of wh-questions in three subtypes of SLI. *Lingua* 121, 367–382. doi:10.1016/j.lingua.2010.10.004
- Friedmann, N., Yachini, M., and Szterman, R. (2015). “Relatively easy relatives: children with syntactic SLI avoid intervention,” in *Structures, Strategies and Beyond. Studies in Honour of Adriana Belletti*, eds E. Di Domenico, C. Hamann, and S. Matteini (Amsterdam: John Benjamins), 303–320.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balcuniene, I., et al. (2015). “Assessment of narrative abilities in bilingual children,” in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 243–276.
- Gallon, N., Harris, J., and van der Lely, H. (2007). Non-word repetition: an investigation of phonological complexity in children with grammatical SLI. *Clin. Linguist. Phon.* 21, 445–455. doi:10.1080/02699200701299982
- Genesee, F., Paradis, J., and Crago, M. (2004). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*. Baltimore: Brookes Publishing.
- Glück, C. (2007). *WWT 6-11. Wortschatz- und Wortfindungstest für 6-10jährige Kinder*. Munich: Elsevier.
- Goldberg, H., Paradis, J., and Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Appl. Psycholinguist.* 29, 41–65. doi:10.1017/S0142716408008003X
- Grimm, A., dos Santos, C., and Chiat, S. (2014). “Can nonwords be language-independent? Cross-linguistic evidence from monolingual and bilingual acquisition of French, German, and Lebanese,” in *Symposium Language Impairment Testing in Multilingual Setting (LITMUS): Disentangling bilingualism and SLI, July 13–19, 2014* (Amsterdam: IASCL).
- Grimm, A., and Hübner, J. (in press). “Nonword repetition by bilingual learners of German: the role of language-specific complexity,” in *Bilingualism and Specific Language Impairment, Bi-SLI 201*, eds C. dos Santos and L. de Almeida (Amsterdam: Benjamins).
- Grimm, A., and Schulz, P. (2014). Specific language impairment and early second language acquisition: the risk of over- and underdiagnosis. *Child Indic. Res.* 7, 821–841. doi:10.1007/s12187-013-9230-6
- Gutiérrez-Clellen, V., and Simon-Cereidjido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English-Speaking children: does the language of assessment matter? *Learn. Disabil.* 25, 48–58. doi:10.1111/j.1540-5826.2009.00300.x
- Håkansson, G., and Nettelbladt, U. (1996). “Similarities between SLI and L2 children. Evidence from the acquisition of Swedish word order,” in *Children's Language*, Vol. 9, eds C. Johnson and J. Gilbert (Mahwah, NJ: Gardner Press), 135–151.
- Hamann, C. (2012). “Bilingual development and language assessment,” in *Proceedings of BUCLD 36*, eds A. K. Biller, E. Y. Chung, and A. E. Kimball (Somerville, MA: Cascadia Press), 1–28.
- Hamann, C. (2015). “Specific language impairment in German speaking children,” in *Specific Language Impairment. Current Trends in Research*, ed. S. Stavrakaki (Amsterdam: Benjamins), 215–251.
- Hamann, C., Chilla, S., Gagarina, N., and Abed Ibrahim, L. (2017). “Syntactic complexity and bilingualism: how (a)typical bilinguals deal with complex structures,” in *Complexity in Acquisition*, ed. E. Di Domenico (New Castle: Cambridge Scholars Publishing), 142–178.
- Hamann, C., Chilla, S., Ruigendijk, E., and Abed Ibrahim, L. (2013). “A German sentence repetition task: testing bilingual Russian/German children,” in *Poster Presented at the COST Meeting in Krakow*, Oldenburg.
- Hamann, C., Penner, Z., and Lindner, K. (1998). German impaired grammar: the clause structure revisited. *Lang. Acquis.* 7, 193–246. doi:10.1207/s15327817la0702-4\_5
- Hamann, C., and Tuller, L. (2014). Genuine versus superficial relatives in French: the depth of embedding factor. *Revisita di Grammatica Generativa* 36, 146–181.
- IBM SPSS 22. (2013). *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.
- Jakubowicz, C., Nash, L., Rigaut, C., and Gérard, C.-L. (1998). Determiners and clitic pronouns in French-speaking children with SLI. *Lang. Acquis.* 7, 113–160. doi:10.1207/s15327817la0702-4\_3
- Kay, J., Lesser, R., and Coltheart, M. (1996). *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia*. Hove: Lawrence Erlbaum Associates.
- Köpke, B., Schmid, M. S., Keijzer, B., and Weilemar, L. (2004). “First language attrition: the next phase,” in *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, eds M. Schmid and S. Köpke (Amsterdam: John Benjamins), 1–43.
- Lein, T., Hamann, C., Rothweiler, M., Abed Ibrahim, L., Chilla, S., and San, H. (2016). “SLI in bilinguals: testing complex syntax and semantics in German,” in *GASLA Proceedings 2015*, eds D. Stringer, J. Garrett, B. Halloran, and S. Mossman (Somerville, MA: Cascadia Proceedings), 124–135.
- Lein, T., Rothweiler, M., and Hamann, C. (2017). “Factors affecting the performance in child heritage Portuguese in Germany,” in *Proceedings of GALA 2015*, ed. H. Demirdache (Newcastle: Cambridge Scholars Publishing), 146–175.
- Leonard, L. (1998). *Children with Specific Language Impairment*. Cambridge: MIT Press.
- Leonard, L. (2014). *Children with Specific Language Impairment*, 2nd Edn. Cambridge: MIT Press.

- MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (California: University of California Press), 281–297.
- Maddieson, I. (2006). Correlating phonological complexity: data and validation. *Linguis. Typol.* 10, 106–123. doi:10.1515/LINGTY.2006.017
- Maddieson, I., Flavier, S., Marsico, E., and Pellegrino, F. (2011). *LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0*. Available at: <http://www.lapsyd.dclish-lyon.cnrs.fr/>
- Marinis, T., and Armon-Lotem, S. (2015). "Sentence repetition," in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 95–124.
- Marinis, T., Armon-Lotem, S., and Pontikas, G. (2017). "Language impairment in bilingual children – state of the art 2017," in *Linguistic Approaches to Bilingualism*, Vol. 7, eds J. Rothman and S. Unsworth (Amsterdam: John Benjamins), 265–276.
- Marshall, C. R., and van der Lely, H. (2009). Effects of word position and stress on onset cluster production: evidence from typical development, SLI and dyslexia. *Language* 85, 39–57. doi:10.1353/lan.0.0081
- Meisel, J. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft* 28, 5–34. doi:10.1515/ZFSW.2009.002
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism. Re-Examining the Age Factor*. Amsterdam: Benjamins.
- Montrul, S. (2010). Current issues in heritage language acquisition. *Annu. Rev. Appl. Linguis.* 30, 3–23. doi:10.1017/S0267190510000103
- Paradis, J. (2010). The interface between bilingual development and specific language impairment. Keynote article for special issue with peer commentaries. *Appl. Psycholinguist.* 31, 3–28. doi:10.1017/S0142716409990373
- Paradis, J. (2011). Individual differences in child English second-language acquisition: comparing child-internal and child-external factors. *Linguis. Approaches Biling.* 1, 213–237. doi:10.1044/jslhr.4304.834
- Paradis, J., Emmerzael, K., and Sorenson Duncan, T. (2010). Assessment of English language learners: using parent report on first language development. *J. Commun. Disord.* 43, 474–497. doi:10.1016/j.jcomdis.2010.01.002
- Paradis, J., Martha, C., and Genesee, F. (2003). "Object clitics as clinical markers of SLI in French: evidence from French-English bilingual children," in *Proceedings of the 27th Annual Boston University Conference on Language Development*, eds B. Beachley, A. Brown, and F. Conlin (Somerville: Cascadia Press), 638–649.
- Paradis, J., Tulpur, Y., and Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: individual variation in long-term outcomes. *J. Child Lang.* 43, 553–580. doi:10.1017/S0305000915000562
- Plante, E., and Vance, R. (1994). Diagnostic accuracy of two tests of preschool language. *Am. J. Speech Lang. Pathol.* 4, 70–76. doi:10.1044/1058-0360.0402.70
- Poljšenská, K., Chiat, S., and Roy, P. (2014). Sentence repetition: what does the task measure? *Int. J. Lang. Commun. Disord.* 50, 106–118. doi:10.1111/1460-6984.12126
- Rice, M., and Wexler, K. (1996). "Tense over time: the persistence of optional infinitives in English in children with SLI," in *Proceedings of the 20th Annual Boston University Conference on Language Development*, eds A. Stringfellow, D. Cahana-Amitay, E. Hughes, and A. Zukowski (Somerville: Cascadia Press), 610–621.
- Rinke, E., and Flores, C. (2014). Morpho-syntactic knowledge of clitics by Portuguese heritage children. *Biling. Lang. Cognit.* 17, 681–699. doi:10.1017/S136672891300076X
- Rizzi, L. (2004). "Locality and left periphery," in *Structures and Beyond. The Cartography of Syntactic Structures*, Vol. 3, ed. A. Belletti (Oxford: OUP), 223–251.
- Rothweiler, M., Chilla, S., and Clahsen, H. (2012). Subject verb agreement in specific language impairment: a study of monolingual and bilingual German-speaking children. *Biling. Lang. Cognit.* 15, 39–57. doi:10.1017/S136672891100037X
- Schönenberger, M., Rothweiler, M., and Sterner, F. (2012). "Case marking in child L1 and early child L2 German," in *Multilingual Individuals and Multilingual Societies. Hamburg Studies on Multilingualism*, eds K. Braunmüller and C. Gabriel (Amsterdam: Benjamins), 3–22.
- Schroeder, C., and Dollnick, M. (2013). "Mehrsprachige Gymnasiasten mit türkischem Hintergrund schreiben auf Türkisch," in *Mehrsprachig in Wissenschaft und Gesellschaft. Tagungsband*, eds C. Riemer, H. Brandl, E. Arslan, and E. Langelahn (Open-Access Publication Publikationsplattform BieColl of the University of Bielefeld), 101–114. doi:10.2390/biecoll-mehrspr2013\_0
- Schulz, P., and Roeper, T. (2011). Acquisition of exhaustivity in wh-questions: a Semantic dimension of SLI? *Lingua*. 121, 383–407. doi:10.1016/j.lingua.2010.10.005
- Schulz, P., and Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe Verlag.
- Sua-Kay, E., and Santos, M. E. (2014). *Grelha de Avaliacao da Linguagem – nível escolar (GOL-E)*, 2nd edição revista Edn. Alcoitao: Escola Superior de Saúde do Alcoitao.
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 14, 426–445. doi:10.1177/1367006911403202
- Thordardottir, E. (2012). *Proposed Diagnostic Procedures and Criteria for Bi-SLI COST Action Studies. COST Action IS080*. Available at: <http://www.bi-sli.org>
- Thordardottir, E. (2015). "Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts," in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 331–358.
- Thordardottir, E., and Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J. Commun. Disord.* 46, 1–16. doi:10.1016/j.jcomdis.2012.08.002
- Tomblin, B. J., Records, N. L., Buckwalter, P. R., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *JSLHR* 40, 1245–1260.
- Topbaş, S., and Güven, S. (2013). *Turkce Erken Dil Gelism Testi (TEDİL) (Test of Early Language Development: TELD-3, Turkish Version)*. Ankara: Detay Yayıncılık.
- Tsimpli, I. M., and Dimitrakopoulou, M. (2007). The interpretability hypothesis: evidence from wh-interrogatives in second language acquisition. *Second Lang. Res.* 23, 215–242. doi:10.1177/0267658307076546
- Tuller, L. (2015). "Clinical use of parental questionnaires in multilingual contexts," in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 229–328.
- Tuller, L., Abboud, L., Ferré, S., Fleckstein, A., Prévost, P., dos Santos, C., et al. (2015). "Specific language impairment and bilingualism: assembling the pieces," in *Language Acquisition and Development. Proceedings of GALA 2013*, eds C. Hamann and E. Ruigendijk (Newcastle: Cambridge Scholars Press), 533–567.
- van der Lely, H. (1998). SLI in children: movement, economy, and deficits in the computational-syntactic system. *Lang. Acquis.* 7, 161–192. doi:10.1027/s15327817a0702-4\_4
- Zebib, R., Henri, G., Khamsi, A., Messara, C., and Hreich, E. (2017). *Batterie d'Evaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Liban: LTE.

**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 © 2017 Hamann and Abed Ibrahim. 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 Working Memory Is Related to the Acquisition of Cross-Linguistic Phonological Regularities

Evelyn Bosma<sup>1,2,3\*</sup>, Wilbert Heeringa<sup>1</sup>, Eric Hoekstra<sup>1</sup>, Arjen Versloot<sup>1,4</sup> and Elma Blom<sup>5</sup>

<sup>1</sup> Fryske Akademy, Leeuwarden, Netherlands, <sup>2</sup> Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, Netherlands, <sup>3</sup> Leiden University Centre for Linguistics, Leiden University, Leiden, Netherlands, <sup>4</sup> Department of Modern Foreign Languages and Cultures, University of Amsterdam, Amsterdam, Netherlands, <sup>5</sup> Special Education Cognitive and Motor Disabilities, Department of Education and Pedagogy, Utrecht University, Utrecht, Netherlands

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Jing Zhao,  
Capital Normal University, China  
Josh S. Payne,  
Bangor University, United Kingdom

### \*Correspondence:

Evelyn Bosma  
e.bosma@hum.leidenuniv.nl

### Specialty section:

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

Received: 20 April 2017

Accepted: 17 August 2017

Published: 12 September 2017

### Citation:

Bosma E, Heeringa W, Hoekstra E,  
Versloot A and Blom E (2017) Verbal  
Working Memory Is Related to the  
Acquisition of Cross-Linguistic  
Phonological Regularities.  
Front. Psychol. 8:1487.  
doi: 10.3389/fpsyg.2017.01487

Closely related languages share cross-linguistic phonological regularities, such as Frisian *-âld* [ɔ:t] and Dutch *-oud* [ɔut], as in the cognate pairs *kâld* [kɔ:t] – *koud* [kɔut] ‘cold’ and *wâld* [wɔ:t] – *woud* [wɔut] ‘forest’. Within Bybee’s (1995, 2001, 2008, 2010) network model, these regularities are, just like grammatical rules within a language, generalizations that emerge from schemas of phonologically and semantically related words. Previous research has shown that verbal working memory is related to the acquisition of grammar, but not vocabulary. This suggests that verbal working memory supports the acquisition of linguistic regularities. In order to test this hypothesis we investigated whether verbal working memory is also related to the acquisition of cross-linguistic phonological regularities. For three consecutive years, 5- to 8-year-old Frisian-Dutch bilingual children ( $n = 120$ ) were tested annually on verbal working memory and a Frisian receptive vocabulary task that comprised four cognate categories: (1) identical cognates, (2) non-identical cognates that either do or (3) do not exhibit a phonological regularity between Frisian and Dutch, and (4) non-cognates. The results showed that verbal working memory had a significantly stronger effect on cognate category (2) than on the other three cognate categories. This suggests that verbal working memory is related to the acquisition of cross-linguistic phonological regularities. More generally, it confirms the hypothesis that verbal working memory plays a role in the acquisition of linguistic regularities.

**Keywords:** bilingualism, cognates, verbal working memory, cross-linguistic phonological regularities, minority language

## INTRODUCTION

Closely related languages such as Frisian and Dutch share cross-linguistic phonological regularities (Sjölin, 1976; Rys, 2009; Taeldeman, 2013). These regularities connect a fixed sequence of phonemes in one language to another fixed sequence of phonemes in the other language. An example of such a regularity is Frisian *-âld* [ɔ:t] and Dutch *-oud* [ɔut], as in the cognate pairs *kâld* [kɔ:t] – *koud* [kɔut] ‘cold’ and *wâld* [wɔ:t] – *woud* [wɔut] ‘forest’. However, not all cognate pairs follow a cross-linguistic regularity. For example, it is not the case that Frisian *a-* [a] as in *amer* [amər] always corresponds to Dutch *e-* [ɛ] as in *emmer* [ɛmər] ‘bucket’. It is thought that

bilingual speakers make use of cross-linguistic phonological regularities to relate the vocabulary of one language to the other and to quickly switch between languages (Sjölin, 1976; Rys, 2009; Taeldeman, 2013). However, as far as we know, there is no psycholinguistic evidence for this claim. Recent research, though, suggests that cross-linguistic phonological regularities do have a mental reality, as children seem to start using them as they grow older (Bosma et al., 2016).

In the present study, we investigated whether the acquisition of cross-linguistic phonological regularities is related to verbal working memory. This could not only give us more insight into the acquisition of these regularities themselves. As we will explain, it may also shed more light on the mechanisms that support language acquisition in general. In what follows, we will first describe our previous study (Bosma et al., 2016) in more detail, followed by a description of how the acquisition of cross-linguistic phonological regularities could be explained within Bybee's (1995, 2001, 2008, 2010) usage-based network model. It was not our intention to test this model or to make theoretical statements. Rather, we used the model as a framework to describe and interpret regularities within the lexicon in a comprehensible way. Within the network model, applied to a bilingual learning context, phonological regularities across languages are similar to grammatical rules within a language. As the acquisition of grammar, but not vocabulary is supported by verbal working memory (Gottardo et al., 1996; McDonald, 2008; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016), this suggests that verbal working memory supports the acquisition of linguistic regularities. If this is the case, then we would expect verbal working memory to be related to the acquisition of cognates with a cross-linguistic phonological regularity, but not to the acquisition of other types of cognates and non-cognates.

In a longitudinal study with three consecutive annual measurements, Bosma et al. (2016) tested 5- to 8-year-old Frisian-Dutch bilingual children on a Frisian receptive vocabulary task that comprised four cognate categories: (1) identical cognates, (2) non-identical cognates with a simple cross-linguistic phonological regularity (3) non-identical cognates without or with a more complex cross-linguistic phonological regularity, and (4) non-cognates. The results showed a gradual cognate facilitation effect for children with a low intensity of exposure to Frisian at home: the higher the degree of cross-language similarity, the better their performance. Furthermore, over time, the children with a low intensity of exposure to Frisian at home improved the most on non-identical cognates with a cross-linguistic phonological regularity. In the first and second year of the study, their performance on this type of cognates was comparable to their performance on non-identical cognates without such a regularity, whereas in the third year of the study, it was similar to their performance on identical cognates. This suggests that as they grow older, children become better at recognizing regularities between the Frisian and Dutch phonological systems.

The graduality of the cognate facilitation effect shows that a word in the input co-activates semantically and phonologically similar words in the other language depending on their degree of similarity. In fact, the spreading of activation in the bilingual

lexicon is probably no different from the spreading of activation in the monolingual lexicon (Costa et al., 2005), which has also been shown to depend on the degree of phonological and semantic similarity between words (Gonnerman et al., 2007). This spreading of lexical activation as a function of similarity is the basis of Bybee's (1995, 2001, 2008, 2010) network model, which proposes that the lexicon is a complex network of linguistic items in which phonologically and semantically related words are stored as spatially proximate. In this model, it is argued that similarity-based categorization and analogy are two of the domain-general mechanisms that support language acquisition. As speakers categorize linguistic items for storage, so-called schemas arise. These are organizational patterns in the lexicon that capture phonological and semantic generalizations about linguistic items. For example, English past tense verbs with the allomorph /d/ are stored together because they have the same final consonant and share past-tense meaning. The connections between these past tense forms lead to the identification of the suffix. When a speaker creates novel items based on analogy to this schema, the past tense suffix becomes productive. In contrast to what is traditionally thought of as grammar, the generalizations that arise from schemas in the lexicon do not necessarily have a cognitive representation that is independent of the individual linguistic items that together form the schema. This means that there is no separate storage of the rule. Within Bybee's network model, grammar is not seen as a system that is separate from the lexicon [as in Pinker's (1991) dual-processing model or Ullman's (2004) declarative/procedural model], but rather as the structure that arises from the complex network of phonological and semantic relations within the lexicon.

As similarity-based activation of lexical items occurs both within (Gonnerman et al., 2007) and across languages (Dijkstra et al., 2010; Bosma et al., 2016), it can be assumed that phonologically and semantically similar words are stored closely together, regardless of whether they belong to the same or to a different language. Thus, the network model is not only able to account for regularities within a language, but also for regularities across languages. This suggests that cross-linguistic phonological regularities resemble grammatical rules, as they can both be thought of as generalizations that arise from schemas of phonologically and semantically related words.

Previous research has shown that grammar acquisition is related to verbal working memory (Gottardo et al., 1996; McDonald, 2008; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016). The precise cognitive architecture of the verbal working memory system is still under debate, but although different researchers work with different definitions (for an overview, see Cowan, 2016), most views support that it is used for both the temporary storage, also referred to as verbal short-term memory, and the processing of verbal information. Following Baddeley and Hitch (1974) and Baddeley (1986), verbal short-term is thus considered to be part of the larger verbal working memory system. Verbal short-term memory has been shown to play a role in children's first (L1) (Gathercole et al., 1992; Gathercole et al., 1997; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016) and second language (L2)

vocabulary acquisition (Cheung, 1996; Masoura and Gathercole, 2005; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016) as well as in children's L1 (Montgomery, 1995) and L2 grammar acquisition (French and O'Brien, 2008; Verhagen et al., 2015; Verhagen and Leseman, 2016). The processing component of verbal working memory is also argued to be important for children's L1 (Gottardo et al., 1996; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016) and L2 grammar acquisition (McDonald, 2008; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016), as has been shown by studies involving receptive grammar (Engel de Abreu and Gathercole, 2012), sentence repetition (Verhagen and Leseman, 2016), grammaticality judgment (Gottardo et al., 1996; McDonald, 2008) and inflectional morphology (Verhagen and Leseman, 2016). However, no relationship has been found between verbal working memory and vocabulary acquisition (Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016). This suggests that verbal short-term memory and verbal working memory are differentially associated with language learning. As both vocabulary and grammar are related to verbal short-term memory, it is argued that the storage component of verbal working memory is important for the development of stable phonological representations in long-term memory (Baddeley et al., 1998). After all, children can only transfer words and multiword units to long-term memory after they have first stored them in short-term memory (Speidel, 1993).

The observation that verbal working memory is related to the acquisition of grammar, but not vocabulary suggests that verbal working memory is important for the processing of linguistic regularities. In terms of Bybee's network model, this suggests that it plays a role in the formation of linguistic schemas through categorization and/or their productive use through analogy, a view that is supported by the finding that verbal working memory also plays a role in the categorization of non-linguistic items (Lewandowsky, 2011; Lewandowsky et al., 2012) and in non-linguistic analogical reasoning (Waltz et al., 2000).

In the current study, we investigated the hypothesis that verbal working memory is related to the acquisition of linguistic regularities. Although previous studies did not find a relationship between verbal working memory and the acquisition of vocabulary (Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016), we expected to find this relationship when the words follow a particular pattern. To this end, we investigated children's vocabulary acquisition in a bilingual context with two closely related languages that share cross-linguistic phonological regularities. We hypothesized that verbal working memory would support the acquisition of cognates that follow a cross-linguistic phonological regularity, but not the acquisition of other types of cognates and non-cognates. In order to answer this question, we used the longitudinal data from the 5- to 8-year-old children in our previous cognate study (Bosma et al., 2016) and investigated associations with verbal working memory, thereby controlling for verbal short-term memory (Engel de Abreu and Gathercole, 2012), SES (Rice and Hoffman, 2015), exposure (Pearson et al., 1997), non-verbal IQ (Rice and Hoffman, 2015) and age, which have previously been shown to be related to vocabulary learning.

## MATERIALS AND METHODS

### Participants

Participants were recruited by contacting primary schools in the countryside of the Dutch province of Fryslân. A total of 122 children from 14 different schools took part in the first year of our study (61 girls and 61 boys). Two children dropped out after the first wave of data collection, leaving 120 children in the second and third year of the study (61 girls and 59 boys). They were 5- or 6-years-old at time 1, 6- or 7-years-old at time 2 and 7- or 8-years-old at time 3. **Table 1** provides an overview of participants' age, non-verbal IQ scores, socioeconomic status (SES) and intensity of exposure to Frisian at home. Non-verbal IQ was measured with the subsets Matrices and Recognition of the *Wechsler Non-verbal Scale of Ability* (WNV; Wechsler and Naglieri, 2006). Information about SES and intensity of exposure to Frisian at home were obtained through a parental questionnaire, based on the *Questionnaire for Parents of Bilingual Children* (PaBiQ) (Cost Action ISO804, 2011; Tuller, 2015). SES was calculated as the mean educational level of the father and the mother of the child, which was measured on a 1 to 9 scale, ranging from no education (1) to university degree (9). Intensity of exposure to Frisian was measured as the mean percentage of Frisian input the child received from his mother, father, siblings and other adults who looked after the child at least once per week. For each of these people the question had to be answered how often (s)he spoke Frisian to the child: 'never' (0%), 'seldom' (25%), 'sometimes' (50%), 'usually' (75%) and 'always' (100%). Intensity of exposure to Dutch at home was 100% minus intensity of exposure to Frisian at home. As SES and IQ (Rice and Hoffman, 2015) and exposure (Pearson et al., 1997) have been shown to be related to vocabulary learning we included these as control variables.

### Measurement Instruments

#### Frisian Receptive Vocabulary

Frisian receptive vocabulary was measured with a task that was based on the Peabody Picture Vocabulary Test-III-NL (PPVT-III-NL; Schlichting, 2005), which is the Dutch version of the PPVT-III (Dunn and Dunn, 1997). Permission was obtained from the publisher to use this Frisian adaptation for research purposes. In this Frisian adaptation [see Bosma et al. (2016) for more details], only the first 144 words of the Dutch PPVT were used. These items suffice to test the vocabulary knowledge of

**TABLE 1** | Descriptive characteristics of the participants.

|               | Mean (SD) (n = 120) | Range  | Maximum possible score |
|---------------|---------------------|--------|------------------------|
| Age at time 1 | 70 (7)              | 59–83  |                        |
| Age at time 2 | 82 (7)              | 71–95  |                        |
| Age at time 3 | 94 (7)              | 83–107 |                        |
| IQ            | 106 (15)            | 73–144 | 144                    |
| SES           | 6.9 (1.3)           | 3.5–9  | 9                      |
| % FR          | 63 (29)             | 0–100  | 100                    |

Age, age in months; IQ, intelligence quotient; SES, socioeconomic status; % FR, intensity of exposure to Frisian at home.

the children in our age range. To make sure that all children completed all items, we did not use basal and ceiling criteria.

Words were assigned to four different cognate categories that differed with respect to degree of cross-language similarity: (1) identical cognates, such as Frisian *poes* [pus] and Dutch *poes* [pus] ‘cat’, (2) non-identical cognates that exhibit a simple phonological regularity between Frisian and Dutch, such as *wâld* [wɔ:t] – *woud* [waut] ‘forest’, (3) non-identical cognates that do not exhibit a simple phonological regularity between Frisian and Dutch, such as Frisian *amer* [amər] and Dutch *emmer* [ɛmər] ‘bucket’ and (4) non-cognates, such as Frisian *bern* [bɛ:n] and Dutch *kind* [kɪnt] ‘child’.

Category (2) comprised items that exhibit a regularity of one, two or three phonemes. An overview of all cross-linguistic phonological regularities of category 2 and some examples can be found in **Table 2**. The vast majority of the items in category (3) were cognates without a cross-linguistic regularity (34 items). Two items followed a more complex cross-linguistic regularity that involves four phonemes. In order to check if the outcomes, in particular differences between category 2 and category 3, were affected by these two items, analyses were run both with and without these items.

As a consequence of how we defined the cognate categories, there was a significant difference between the four categories regarding the number of phoneme differences between the Frisian and Dutch translation equivalents.  $F(3,140) = 93.47$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.67$  (category 1:  $M = 0.00$ ,  $SD = 0.00$ ; category 2:  $M = 1.86$ ,  $SD = 0.99$ ; category 3:  $M = 2.92$ ,  $SD = 1.25$ ; category 4:  $M = 5.72$ ,  $SD = 2.50$ ). Pairwise comparisons showed that all differences between categories were significant at the  $p < 0.01$  level. There were, however, no significant differences between the four cognate categories with respect to the number of phonemes per word,  $F(3,140) = 0.95$ ,  $p = 0.42$ ,  $\eta_p^2 = 0.02$  (category 1:  $M = 6.17$ ,  $SD = 2.06$ ; category 2:  $M = 6.75$ ,  $SD = 2.63$ ; category 3:  $M = 5.83$ ,  $SD = 2.04$ ; category 4:  $M = 6.17$ ,  $SD = 2.60$ ).

Furthermore, it was ensured that there were no word frequency differences between the four categories. The only available corpus for Frisian is a non-lemmatized database of

standardized written language, which is not representative of the language that is spoken by speakers of Frisian (Breuker, 1993). Therefore, we used frequencies per million words from two Dutch corpora instead: CELEX (Center for Lexical Information, 1993), which is a corpus of written Dutch that was also used for the PPVT-III-NL, and *Corpus Gesproken Nederlands* (“Corpus Spoken Dutch”; CGN; Nederlandse Taalunie, 2004), which is a corpus of spoken Dutch. As Frisian and Dutch are closely related languages, the Dutch frequencies were thought to be representative of the Frisian frequencies. As frequency is perceived logarithmically, we calculated Zipf scores (Van Heuven et al., 2014), which are based on logarithmic (10-log) instead of absolute frequencies.

The four cognate categories each had about the same frequencies in CELEX and CGN, which was also confirmed by the high correlation between the CELEX and the CGN frequencies,  $r = 0.75$ ,  $p < 0.001$ . A One-Way ANOVA with category as the independent variable and CELEX frequencies as the dependent variable showed that there was no significant effect of CELEX frequency,  $F(3,140) = 0.24$ ,  $p = 0.87$ , and that the CELEX frequencies of category 1 ( $M = 3.82$ ,  $SD = 0.92$ ), category 2 ( $M = 3.85$ ,  $SD = 1.39$ ), category 3 ( $M = 4.04$ ,  $SD = 1.22$ ) and category 4 ( $M = 3.96$ ,  $SD = 1.37$ ) could be assumed to be the same. A One-Way ANOVA with category as the independent variable and CGN frequencies as the dependent variable showed that there was also no significant effect of CGN frequency,  $F(3,140) = 0.40$ ,  $p = 0.76$ , and that the CGN frequencies of category 1 ( $M = 3.71$ ,  $SD = 0.66$ ), category 2 ( $M = 3.79$ ,  $SD = 0.86$ ), category 3 ( $M = 3.93$ ,  $SD = 1.05$ ) and category 4 ( $M = 3.85$ ,  $SD = 0.99$ ) could be assumed to be the same. Furthermore, Cronbach’s alpha, as calculated at time 1, showed that the internal consistency of the items in the test was sufficient,  $\alpha = 0.76$ .

## Verbal Memory

Both verbal short-term memory and verbal working memory were measured, as this allowed us to separate the storage component of verbal working memory from the processing

**TABLE 2 |** Cross-linguistic phonological regularities category 2.

| Frisian phoneme(s) | Frisian example         | Dutch phoneme(s) | Dutch example             | English translation |
|--------------------|-------------------------|------------------|---------------------------|---------------------|
| [u:]               | klūs [klu:s]            | [œy]             | kluis [kloëys]            | safe                |
| [u]                | pûlfrucht [pulfrøxt]    | [ø:]             | peulvrucht [pø:lvrøxt]    | legume              |
|                    | ûnder [undər]           | [o]              | onder [ondər]             | under               |
| [sk]               | skep [skɛp]             | [sx]             | schep [sxɛp]              | shovel              |
| [ɔ:n]              | hân [hɔ:n]              | [ant]            | hand [hant]               | hand                |
| [ɔ:t]              | kâld [kɔ:t]             | [aut]            | koud [kaut]               | cold                |
| [a:]               | daam [da:m]             | [ɑ]              | dam [dam]                 | dam                 |
| [ər]               | ferstelber [fɛrstɛlbər] | [a:r]            | verstelbaar [vɛrstɛlba:r] | adjustable          |
| [l.ə]              | easten [l.ɛstən]        | [o:]             | oosten [o:stən]           | east                |
| [i]                | dolfyn [dɔlfɪn]         | [ɛi]             | dolfijn [dɔlfɛin]         | dolphin             |
| [(k)jə]            | timmerje [tɪmərjə]      | [ən]             | timmeren [tɪmərən]        | to hammer           |
| [jɛrjə]            | dosearje [do:sjɛrjə]    | [l.ərən]         | doceren [do:sl.ərən]      | to teach            |
| [tsjə]             | kadootsje [kado:tsjə]   | [tjə]            | cadeautje [kado:tjə]      | (little) present    |
| [kə]               | groepke [grupkə]        | [jə]             | groepje [xrupjə]          | (small) group       |
|                    | boeid [buit]            | [xə]             | geboeid [xəbuit]          | chained             |

component. Verbal short-term memory was measured with the Forward Digit Span and verbal working memory with the Backward Digit Span. These tasks were based on the Alloway Working Memory Assessment (AWMA; Alloway, 2012) and translated to Dutch. It was assumed that all children were able to count to 10 in Dutch, since Dutch is the main language of education and all children had spent at least 1 year in education at the first time of testing. In the forward version of the task, children had to repeat sequences of digits in the same order, whereas in the Backward Digit Span, they had to repeat them in reversed order. The Forward Digit Span is considered a measure of verbal short-term memory, because it only requires the storage of the digits. The Backward Digit Span, in contrast, is considered a measure of verbal working memory, because the added requirement to recall the digits in reversed order imposes a substantial processing load on the child (Alloway et al., 2008).

The task started with sequences of one digit, after which the sequences became increasingly longer. Per block, there were six trials and after three incorrect trials within one block the task stopped. When the child repeated the first four trials within one block correctly, he or she automatically continued with the next block and received a score of six. When the child repeated four out of the first five trials correctly, he or she also automatically continued with the next block and received a score of five. The AWMA procedure (Alloway, 2012) was applied for scoring. Trials were scored as incorrect if (part of) the sequence was incorrect, if children recalled one or more digits incorrectly, or if they omitted one or more digits. There were seven blocks for both the Forward and the Backward Digit Span, so the scores could range from 0 to 42.

## Procedure

The schools distributed consent forms and folders providing information about the experiment among the parents of the children. Children whose parents had signed the consent form were tested individually in a quiet room at school, except for one child at time 1, four children at time 2 and five children at time 3, who were tested at home. The children were tested by the first author and two research assistants, who all had a native level command of both Frisian and Dutch. The tasks

in this study were part of a larger test battery that included language and cognitive tasks that are not reported on in the current study. Children were tested on all tasks at all three time points.

## RESULTS

### Descriptive Statistics

Means and standard deviations for the Forward Digit Span, the Backward Digit Span and the four cognate categories are given in **Table 3**. Repeated measures ANOVAs showed that over time, children improved on all measures,  $p < 0.001$ . Bivariate correlations among all variables at time 1, 2 and 3 are reported in the **Appendix**.

### Mixed Models Analysis

The research question of the current study was whether verbal working memory is related to the acquisition of cross-linguistic phonological regularities. We investigated this research question by examining whether the Backward Digit Span (verbal working memory) had a stronger effect on vocabulary items from cognate category (2) than on vocabulary items from cognate category (1), (3) and (4). In order to answer the research question we used a cumulative link mixed model. The mixed model was run, using the `clmm` function as implemented in the R package `ordinal` (Christensen, 2015). We entered Frisian receptive vocabulary accuracy as the ordered dependent variable, with 1 indicating a correct answer and 0 indicating an incorrect answer. We included random intercepts for subject and item, as both of these variables had repeated values. Including random intercepts would allow us to generalize the outcomes to the larger population of Frisian-Dutch bilingual children and to other items. A manual stepwise model selection procedure was carried out in which factors were added in such a way that the Akaike Information Criterion (AIC) was minimized. This procedure was applied with Category and Backward Digit Span as the main predictors of our study. In addition, the following predictors were added as control variables: Time, Frisian exposure at home, SES, IQ, Age and Forward Digit Span (verbal short-term memory). Time was added as an ordered factor, with  $1 < 2 < 3$ . All of the predictors, except for Category, improved the model fit and were thus included in

**TABLE 3 |** Means and standard deviations for the Forward Digit Span, the Backward Digit Span and the four cognate categories.

|                           | Time 1 <i>M (SD)</i> | Time 2 <i>M (SD)</i> | Time 3 <i>M (SD)</i> | <i>p</i> | $\eta^2$ |
|---------------------------|----------------------|----------------------|----------------------|----------|----------|
| <b>Memory measures</b>    |                      |                      |                      |          |          |
| Forward Digit Span        | 20.11 (3.51)         | 22.47 (3.92)         | 24.11 (3.35)         | <0.001   | 0.534    |
| Backward Digit Span       | 12.75 (2.92)         | 14.90 (2.88)         | 16.47 (3.57)         | <0.001   | 0.391    |
| <b>Cognate categories</b> |                      |                      |                      |          |          |
| Category 1                | 23.11 (2.46)         | 25.18 (2.16)         | 26.23 (2.22)         | <0.001   | 0.440    |
| Category 2                | 22.35 (2.92)         | 24.42 (2.59)         | 26.23 (2.75)         | <0.001   | 0.475    |
| Category 3                | 22.79 (3.17)         | 24.51 (2.73)         | 25.99 (2.49)         | <0.001   | 0.403    |
| Category 4                | 22.03 (4.05)         | 23.87 (3.61)         | 24.78 (2.92)         | <0.001   | 0.279    |

*Category 1 = identical cognates; category 2 = cognates with a simple rule; category 3 = cognates without a simple rule; category 4 = non-cognates.*

the final model. As expected, higher scores on exposure, SES, non-verbal IQ, age and Backward Digit Span were related to better performance on Frisian receptive vocabulary. Time was not a significant predictor, but was added to the final model, as the AIC showed that it did improve the fit. Furthermore, it must be noted that the Forward Digit Span was only significant when the Backward Digit Span was not included in the model.

The model was further refined in an exploratory way by adding potential interactions between the predictors, including Category. This was done in order to increase the amount of explained variance, which would give a better focus on the variables of interest. Interactions between Category and Exposure, Category and Forward Digit Span, and Category and Backward Digit Span significantly improved the model fit and were therefore included in the final model. Models with three-way interactions including Time did not converge. In order to examine the interaction effects in more detail, the model was run four times with different reference levels for Category (1, 2, 3, and 4). We will first discuss the control interactions (Category  $\times$  Exposure, Category  $\times$  Forward Digit Span), followed by the interaction of interest (Category  $\times$  Backward Digit Span). The interaction effect between Category and Exposure showed that the effect of Exposure on Frisian vocabulary was strongest for category (4), followed by category (3), category (2), and category (1) ( $4 > 3 > 2 > 1$ ). The interaction effect between Category and Forward Digit Span showed that the effect of Forward Digit Span on Frisian vocabulary was significantly stronger for items from category (1) than for items from category (3) and (4), and stronger for items from category

(2) than for items from category (4) ( $1 > 3, 4; 2 > 4$ ). This shows that the effect of Forward Digit Span was stronger for items with a high degree of overlap across Frisian and Dutch than for items with a low degree of overlap, although the effect of Forward Digit Span on two adjacent categories was never significantly different. Finally, we examined the interaction effect between Category and Backward Digit Span, which was the focus of the current study. The results showed that the Backward Digit Span had a significantly stronger effect on vocabulary items from category (2) than on vocabulary items from category (1), (3) and (4). The differences between categories (1), (3) and (4) were statistically non-significant ( $2 > 1, 3, 4$ ). The results of the final model are reported in **Table 4**, with category (2) as the reference level, as this category was the focus of our study. **Figure 1** shows the interaction effect between Category and Backward Digit Span. In this figure, it can be seen that the slope of category (2) is steeper than the slope of the other three categories.

As explained in the Method section, there were two items from category (3) that followed a more complex cross-linguistic phonological regularity. In order to check if these items affected the outcomes, the analyses described above were rerun without these items. The results showed that excluding these two items did not affect the outcomes.

We considered that the effect of Time, Exposure, SES, IQ, Age, Forward Digit Span, Backward Digit Span and Category may be different per subject and per item. Therefore, we added several combinations of these variables as random slopes to subject and item. We found that adding Age as random slope to subject and the factors Age, Time and Forward Digit Span

**TABLE 4 |** Fixed effects from the final model with Frisian receptive vocabulary accuracy as dependent variable and category 2 as reference level.

| Effect                                  | Estimate | Std. Error | Z-value | P-value   |
|-----------------------------------------|----------|------------|---------|-----------|
| Time.Linear                             | 0.081    | 0.073      | 1.118   | 0.264     |
| Time.Quadratic                          | -0.043   | 0.022      | -1.941  | 0.052     |
| Category 1                              | 0.316    | 0.610      | 0.517   | 0.605     |
| Category 3                              | -0.210   | 0.605      | -0.347  | 0.729     |
| Category 4                              | -0.520   | 0.605      | -0.859  | 0.390     |
| Frisian exposure                        | 0.079    | 0.035      | 2.279   | 0.023*    |
| SES                                     | 0.100    | 0.026      | 3.778   | <0.001*** |
| Non-verbal IQ                           | 0.085    | 0.028      | 3.056   | 0.002**   |
| Age                                     | 0.280    | 0.049      | 5.730   | <0.001*** |
| Forward Digit Span                      | 0.023    | 0.036      | 0.650   | 0.515424  |
| Backward Digit Span                     | 0.157    | 0.034      | 4.641   | <0.001*** |
| Category 1 $\times$ exposure            | -0.123   | 0.037      | -3.314  | <0.001*** |
| Category 3 $\times$ exposure            | 0.212    | 0.036      | 5.963   | <0.001*** |
| Category 4 $\times$ exposure            | 0.368    | 0.035      | 10.484  | <0.001*** |
| Category 1 $\times$ Forward Digit Span  | 0.066    | 0.044      | 1.490   | 0.136195  |
| Category 3 $\times$ Forward Digit Span  | -0.032   | 0.042      | -0.768  | 0.442396  |
| Category 4 $\times$ Forward Digit Span  | -0.087   | 0.041      | -2.108  | 0.035*    |
| Category 1 $\times$ Backward Digit Span | -0.124   | 0.045      | -2.785  | 0.005**   |
| Category 3 $\times$ Backward Digit Span | -0.151   | 0.042      | -3.563  | <0.001*** |
| Category 4 $\times$ Backward Digit Span | -0.157   | 0.042      | -3.758  | <0.001*** |

Category 1 = identical cognates; category 2 = cognates with a simple rule; category 3 = cognates without a simple rule; category 4 = non-cognates; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

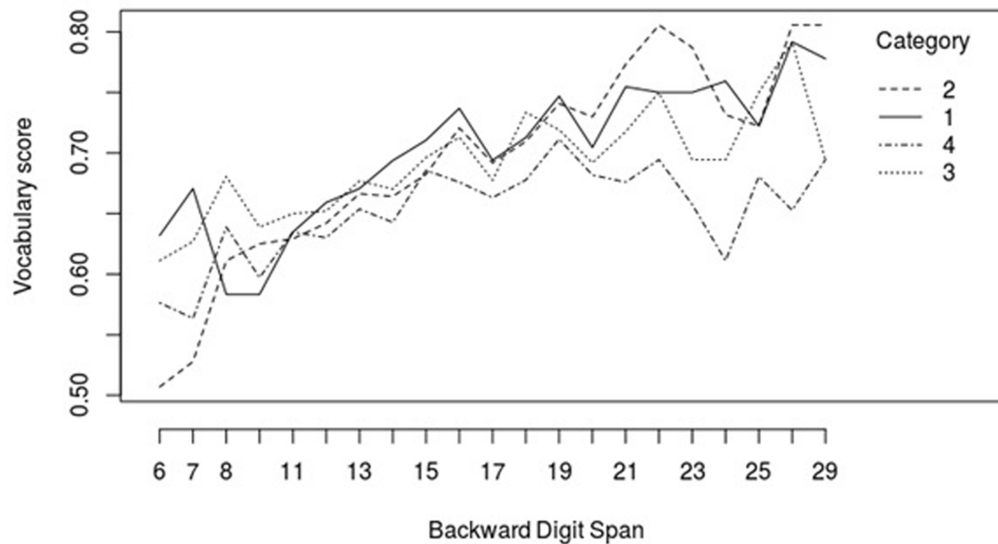


**FIGURE 1 |** Interaction effect between Category and Backward Digit Span on Frisian receptive vocabulary accuracy. Category 1 = identical cognates; category 2 = cognates with a simple rule; category 3 = cognates without a simple rule; category 4 = non-cognates.

as random slopes to item improved the model fit and slightly changed the results, with Time now being a significant predictor,  $p = 0.016$ . However, when we tried to rerun this model with the same random slopes but without the Backward Digit Span as a predictor, the model did not converge. The same problem occurred when we tried to rerun the model with the same random slopes but without the two items from category (3) that followed a more complex cross-linguistic phonological regularity.

## DISCUSSION

Previous research has shown that verbal working memory is related to the acquisition of grammar, but not vocabulary (e.g., Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016). This suggests that verbal working memory supports the acquisition of linguistic regularities. In the present study, we investigated this hypothesis by examining whether verbal working memory is also related to the acquisition of cross-linguistic phonological regularities, such as Frisian *-âld* [ɑ:t] and Dutch *-oud* [ɑut], as in the cognate pairs *kâld* [kɑ:t] – *koud* [kɑut] ‘cold’ and *wâld* [wɑ:t] – *woud* [wɑut] ‘forest’. In order to answer this question, 5- to 8-year-old Frisian-Dutch bilingual children were tested annually for a 3-year period on verbal working memory and a Frisian receptive vocabulary task with four cognate categories: (1) identical cognates, (2) non-identical cognates that either do or (3) do not exhibit a phonological regularity between Frisian and Dutch, and (4) non-cognates. As age, non-verbal IQ (Rice and Hoffman, 2015), exposure (Pearson et al., 1997), SES (Rice and Hoffman, 2015) and verbal short-term memory (Engel de Abreu and Gathercole, 2012) have previously been shown to be related to vocabulary acquisition, these were also measured and included as control variables.

In line with previous studies, the results showed significant main effects of age, SES, non-verbal IQ and exposure on Frisian receptive vocabulary, with higher scores on these variables resulting in better vocabulary scores. Verbal short-term memory was only significant when verbal working memory was not included in the model. When a model was run that included both verbal short-term memory and verbal working memory, only verbal working memory came out as a significant predictor. This is probably due to the fact that, according to some definitions (Baddeley and Hitch, 1974; Baddeley, 1986), verbal short-term memory is part of verbal working memory. In addition to these main effects, we found interaction effects between cognate category and exposure, cognate category and verbal short-term memory, and cognate category and verbal working memory. As the first two interactions were only added as control variables to improve the model, we will not discuss these here, but instead concentrate on the interaction between cognate category and verbal working memory, which was the focus of the current study. The interaction between cognate category and verbal working memory showed that verbal working memory had a significantly stronger effect on cognate category (2) than on cognate category (1), (3) and (4). This suggests that verbal working memory supports the acquisition of regularities across the Frisian and Dutch phonological systems.

The finding that verbal working memory supports the acquisition of cross-linguistic phonological regularities is noteworthy for the following reasons. First, it provides psycholinguistic evidence for the existence of cross-linguistic phonological regularities (Sjölin, 1976; Rys, 2009; Taeldeman, 2013). Second, it confirms that bilingual children learn these regularities (Bosma et al., 2016) by showing that they do so on the basis of a general cognitive capacity, namely verbal

working memory. Third, the results suggest that the acquisition of phonological regularities across languages shares important characteristics with the acquisition of grammatical rules within a language, which has previously been shown to be related to verbal working memory (Gottardo et al., 1996; McDonald, 2008; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016). Fourth, as both the acquisition of grammar and the acquisition of cross-linguistic phonological regularities are related to verbal working memory, this suggests that verbal working memory plays a role in the acquisition of linguistic regularities.

These results can well be explained within the framework of Bybee's (1995, 2001, 2008, 2010) network model, although we do not exclude the possibility that other models may also fit the data. As Costa et al. (2005) already mentioned, the spreading of activation within the bilingual lexicon (Dijkstra et al., 2010; Bosma et al., 2016) is similar to the spreading of activation within the monolingual lexicon (Gonnerman et al., 2007). Within the network model, this implies that related words are stored together, regardless of whether they belong to the same or to a different language. This suggests that the acquisition of phonological regularities across languages shares important characteristics with the acquisition of grammatical relations within a language, as they are both generalizations that emerge from schemas of phonologically and semantically related words. Our finding that the acquisition of cross-linguistic phonological regularities is related to verbal working memory supports this suggestion, as previous research has shown that the acquisition of grammar is also related to verbal working memory (Gottardo et al., 1996; McDonald, 2008; Engel de Abreu and Gathercole, 2012; Verhagen and Leseman, 2016). In terms of Bybee's network model, this parallel between cross-linguistic regularities and grammar suggests that verbal working memory plays a role in the formation of linguistic schemas through categorization and/or their productive use through analogy, a view that is in line with previous evidence that verbal working memory also plays a role in the categorization of non-linguistic items (Lewandowsky, 2011; Lewandowsky et al., 2012) and in non-linguistic analogical reasoning (Waltz et al., 2000).

There are a number of limitations to the present study that are relevant to mention. First, although we only investigated the role of verbal working memory in the acquisition of cross-linguistic phonological regularities, other cognitive skills might play a role as well. An example of another skill that may influence the acquisition of cross-linguistic phonological regularities is phonological awareness, which is the conscious ability to detect and differentiate between the sounds of a word and to manipulate phonemes to create new words. Previous research has shown that phonological awareness positively influences reading and spelling acquisition, because children with high phonological awareness skills are better able to identify and use letter-sound correspondences (Ehri et al., 2001). In the same way, phonological awareness might help children to identify and use correspondences between the phonological systems of two languages.

A second limitation of the current study is that we investigated the acquisition of cross-linguistic phonological regularities in general, without zooming in on differences that might exist between different types of regularities. Within the network model, it is argued that the productivity of a regularity is to a large extent determined by its type frequency, that is, the number of items that follow that regularity. The more items a schema encompasses, the stronger it is, and the higher the likelihood that the pattern will be extended to novel items. Type frequency interacts with degree of schematicity, that is, the degree of dissimilarity of the members of a class. Highly schematic classes include a wide range of dissimilar items. For example, the English past tense has a high degree of schematicity, as it can be applied to all verbs, no matter their phonological form. In the network model, it is argued that a high type frequency in combination with a high degree of schematicity results in a maximally productive construction. For future research, it would be interesting to examine to what extent the acquisition of cross-linguistic phonological regularities depends on type frequency and degree of schematicity and whether type frequency and schematicity interact with verbal working memory.

Taken together, the main finding of this study is that verbal working memory is related to the acquisition of cross-linguistic phonological regularities. This supports the hypothesis that verbal working memory plays a role in the acquisition of linguistic regularities, thus providing more insight into the mechanisms that facilitate language acquisition.

## ETHICS STATEMENT

All the parents of the participating children gave their written informed consent, as was stated in the Section "Materials and Methods" of our paper. Unfortunately, the study was not officially evaluated by an ethics committee before the start of the study due to a miscommunication. In hindsight, the ethics committee of the University of Amsterdam evaluated the information folder and the informed consent form that we used and came to the conclusion that the research had been conducted with the wellbeing of the participants in mind.

## AUTHOR CONTRIBUTIONS

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

## FUNDING

This work was supported by the Province of Fryslân and by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613465.

## REFERENCES

- Alloway, T. P. (2012). *Alloway Working Memory Assessment 2 (AWMA-2)*. London: Pearson.
- Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliott, J. (2008). Evaluating the validity of the automated working memory assessment. *Educ. Psychol.* 28, 725–734. doi: 10.1080/01443410802243828
- Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi: 10.1037/0033-295X.105.1.158
- Baddeley, A. D. (1986). *Working Memory*. Oxford Psychology Series #11. Oxford: Clarendon Press.
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1
- Bosma, E., Blom, E., Hoekstra, E., and Versloot, A. (2016). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary. *Int. J. Biling. Educ. Biling.* 1–15. doi: 10.1080/13670050.2016.1254152
- Breuker, P. (1993). *Noarmaspekten fan it Hoeddeiske Frysk*. Groningen: Stifting FFYRUG.
- Bybee, J. (1995). Regular morphology and the lexicon. *Lang. Cogn. Process.* 10, 425–455. doi: 10.1080/01690969508407111
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511612886
- Bybee, J. (2008). "Usage-based grammar and second language acquisition," in *Handbook of Cognitive Linguistics and Second Language Acquisition*, eds P. Robinson and N. Ellis (New York, NY: Routledge), 216–236.
- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526
- Center for Lexical Information (1993). *CELEX Lexical Database*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Dev. Psychol.* 32, 867–873. doi: 10.1037/0012-1649.32.5.867
- Christensen, R. H. B. (2015). *Ordinal - Regression Models for Ordinal Data*. Available at: <https://cran.r-project.org/web/packages/ordinal/index.html>
- Cost Action ISO804 (2011). *Questionnaire for Parents of Bilingual Children (PaBiQ)*. Available at: <http://www.bi-sli.org>
- Costa, A., Santesteban, M., and Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain Lang.* 94, 94–103. doi: 10.1016/j.bandl.2004.12.002
- Cowan, N. (2016). The many faces of working memory and short-term storage. *Psychon. Bull. Rev.* doi: 10.3758/s13423-016-1191-6 [Epub ahead of print].
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., and Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *J. Mem. Lang.* 62, 284–301. doi: 10.1016/j.jml.2009.12.003
- Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary*, 3rd Edn. Circle Pines, MN: AGS.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., and Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: evidence from the national reading panel's meta-analysis. *Read. Res. Quart.* 36, 250–287. doi: 10.1598/RRQ.36.3.2
- Engel de Abreu, P. M. J., and Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *J. Educ. Psychol.* 104, 974–986. doi: 10.1037/a0028390
- French, L. M., and O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Appl. Psycholing.* 29, 463–487. doi: 10.1017/S0142716408080211
- Gathercole, S. E., Hitch, G. J., and Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Dev. Psychol.* 33, 966–979. doi: 10.1037/0012-1649.33.6.966
- Gathercole, S. E., Willis, C. S., Emslie, H., and Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: a longitudinal study. *Dev. Psychol.* 28, 887–898. doi: 10.1037/0012-1649.28.5.887
- Gonnerman, L. M., Seidenberg, M. S., and Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: evidence for a distributed connectionist approach to morphology. *J. Exp. Psychol.* 136, 323–345. doi: 10.1037/0096-3445.136.2.323
- Gottardo, A., Stanovich, K. E., and Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. *J. Exp. Child Psychol.* 63, 563–582. doi: 10.1006/jecp.1996.0062
- Lewandowsky, S. (2011). Working memory capacity and categorization: individual differences and modeling. *J. Exp. Psychol.* 37, 720–738. doi: 10.1037/a0022639
- Lewandowsky, S., Yang, L. X., Newell, B. R., and Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *J. Exp. Psychol.* 38, 881–904. doi: 10.1037/a0027298
- Masoura, E. V., and Gathercole, S. E. (2005). Phonological short-term memory skills and new word learning in young Greek children. *Memory* 13, 422–429.
- McDonald, J. (2008). Grammaticality judgments in children: the role of age, working memory, and phonological ability. *J. Child Lang.* 35, 247–268. doi: 10.1017/S0305000907008367
- Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: the role of phonological working memory. *J. Speech Lang. Hear. Res.* 38, 187–199. doi: 10.1044/jshr.3801.187
- Nederlandse Taalunie (2004). *Corpus Gesproken Nederlands, Versie 6.1*. The Hague: Nederlandse Taalunie.
- Pearson, B. Z., Fernández, S. C., Lewedeg, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Appl. Psycholinguist.* 18, 41–58. doi: 10.1017/S0142716400009863
- Pinker, S. (1991). Rules of language. *Science* 253, 530–535. doi: 10.1126/science.1857983
- Rice, M. L., and Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: a longitudinal study from 2;6 to 21 years of age. *J. Speech Lang. Hear. Res.* 58, 345–359. doi: 10.1044/2015\_JSLHR-L-14-0150
- Rys, K. (2009). Intersystemic correspondence rules and headwords in Dutch dialect lexicography. *Int. J. Lexicogr.* 22, 129–150. doi: 10.1093/ijl/lec008
- Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Harcourt Test Publishers.
- Sjölin, B. (1976). *"Min Frysk": Een Onderzoek naar het Ontstaan van Transfer en "Code-Switching" in Gesproken Fries*. Groningen: Frysk Ynstittút oan de Ryksuniversiteit.
- Speidel, G. E. (1993). Phonological short-term memory and individual differences in learning to speak: a bilingual case study. *First Lang.* 13, 69–91.
- Taeldeman, J. (2013). "Recent developments in the East Flemish dialect area," in *Language and Space, Dutch*, Vol. 3, eds F. Hinskens and J. Taeldeman (Berlin: De Gruyter Mouton), 255–276.
- Tuller, L. (2015). "Clinical use of parental questionnaires in multilingual contexts," in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 301–330.
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). Subtlex-UK: a new and improved word frequency database for British english. *Q. J. Exp. Psychol.* 67, 1176–1190. doi: 10.1080/17470218.2013.850521
- Verhagen, J., and Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language

- learners. *J. Exp. Child Psychol.* 141, 65–82. doi: 10.1016/j.jecp.2015.06.015
- Verhagen, J., Leleman, P., and Messer, M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Lang. Learn.* 65, 417–448. doi: 10.1111/lang.12101
- Waltz, J. A., Lau, A., Grewal, S. K., and Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Mem. Cogn.* 28, 1205–1212. doi: 10.3758/BF03211821
- Wechsler, D., and Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability (WNV)*. San Antonio, TX: PsychCorp.

**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 © 2017 Bosma, Heeringa, Hoekstra, Versloot and Blom. 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.

## APPENDIX

**TABLE |** Bivariate correlations among all variables at Time 1.

|       | SES    | IQ     | % FR     | FW DS   | BW DS    | cat1     | cat2     | cat3     | cat4     |
|-------|--------|--------|----------|---------|----------|----------|----------|----------|----------|
| Age   | −0.104 | −0.013 | 0.095    | 0.259** | 0.265**  | 0.341*** | 0.405*** | 0.178*   | 0.271**  |
| SES   | –      | 0.062  | −0.253** | 0.145   | 0.047    | 0.092    | −0.052   | 0.035    | −0.107   |
| IQ    |        | –      | −0.014   | 0.208*  | 0.374*** | 0.265**  | 0.371*** | 0.072    | 0.149    |
| % FR  |        |        | –        | 0.002   | 0.039    | −0.034   | 0.244**  | 0.495*** | 0.625*** |
| FW DS |        |        |          | –       | 0.479*** | 0.304*** | 0.265*   | 0.098    | 0.112    |
| BW DS |        |        |          |         | –        | 0.245*   | 0.422*** | 0.080    | 0.160    |
| cat1  |        |        |          |         |          | –        | 0.391*** | 0.171    | 0.233**  |
| cat2  |        |        |          |         |          |          | –        | 0.359*** | 0.493*** |
| cat3  |        |        |          |         |          |          |          | –        | 0.604*** |

*Cat1 = identical cognates; cat2 = cognates with a simple rule; cat3 = cognates without a simple rule; cat4 = non-cognates; % FR = intensity of exposure to Frisian at home; FW DS, Forward Digit Span; BW DS, Backward Digit Span; \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .*

**TABLE |** Bivariate correlations among all variables at Time 2.

|       | SES    | IQ     | % FR     | FW DS  | BW DS    | cat1     | cat2     | cat3     | cat4     |
|-------|--------|--------|----------|--------|----------|----------|----------|----------|----------|
| Age   | −0.115 | −0.026 | 0.100    | 0.074  | 0.126    | 0.301*** | 0.400*** | 0.330*** | 0.159    |
| SES   | –      | 0.039  | −0.244** | 0.129  | 0.141    | 0.089    | 0.184*   | 0.057    | −0.118   |
| IQ    |        | –      | −0.007   | 0.209* | 0.265**  | −0.080   | 0.153    | 0.122    | 0.047    |
| % FR  |        |        | –        | −0.012 | 0.016    | −0.124   | 0.195*   | 0.412*** | 0.620*** |
| FW DS |        |        |          | –      | 0.491*** | 0.158    | 0.154    | 0.059    | −0.045   |
| BW DS |        |        |          |        | –        | 0.116    | 0.284**  | 0.181*   | 0.122    |
| cat1  |        |        |          |        |          | –        | 0.348*** | 0.176    | 0.136    |
| cat2  |        |        |          |        |          |          | –        | 0.526*** | 0.400*** |
| cat3  |        |        |          |        |          |          |          | –        | 0.544*** |

*Cat1 = identical cognates; cat2 = cognates with a simple rule; cat3 = cognates without a simple rule; cat4 = non-cognates; % FR = intensity of exposure to Frisian at home; FW DS, Forward Digit Span; BW DS, Backward Digit Span; \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .*

**TABLE |** Bivariate correlations among all variables at Time 3.

|       | SES    | IQ     | % FR     | FW DS  | BW DS    | cat1     | cat2     | cat3     | cat4     |
|-------|--------|--------|----------|--------|----------|----------|----------|----------|----------|
| age   | −0.123 | −0.028 | 0.099    | 0.124  | 0.102    | 0.277**  | 0.251**  | 0.150    | 0.237**  |
| SES   | –      | 0.039  | −0.244** | 0.173  | 0.151    | 0.071    | 0.251**  | 0.115    | 0.016    |
| IQ    |        | –      | −0.007   | 0.171  | 0.332*** | 0.214*   | 0.249**  | 0.184*   | 0.086    |
| % FR  |        |        | –        | −0.039 | −0.001   | −0.118   | −0.097   | 0.455*** | 0.585*** |
| FW DS |        |        |          | –      | 0.380*** | 0.280**  | 0.268**  | 0.074    | −0.052   |
| BW DS |        |        |          |        | –        | 0.292*** | 0.383*** | 0.147    | 0.016    |
| cat1  |        |        |          |        |          | –        | 0.588*** | 0.188*   | 0.211*   |
| cat2  |        |        |          |        |          |          | –        | 0.340*** | 0.229*   |
| cat3  |        |        |          |        |          |          |          | –        | 0.543*** |

*Cat1 = identical cognates; cat2 = cognates with a simple rule; cat3 = cognates without a simple rule; cat4 = non-cognates; % FR = intensity of exposure to Frisian at home; FW DS, Forward Digit Span; BW DS, Backward Digit Span; \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .*



# How Does L1 and L2 Exposure Impact L1 Performance in Bilingual Children? Evidence from Polish-English Migrants to the United Kingdom

Ewa Haman<sup>1\*</sup>, Zofia Wodniecka<sup>2\*</sup>, Marta Marecka<sup>2</sup>, Jakub Szewczyk<sup>2</sup>, Marta Białecka-Pikul<sup>3</sup>, Agnieszka Otwinowska<sup>4</sup>, Karolina Mieszkowska<sup>1</sup>, Magdalena Łuniewska<sup>1</sup>, Joanna Kołak<sup>1</sup>, Aneta Miękiś<sup>1</sup>, Agnieszka Kacprzak<sup>1</sup>, Natalia Banasik<sup>1</sup> and Małgorzata Foryś-Nogala<sup>1</sup>

## OPEN ACCESS

### Edited by:

Maria Teresa Guasti,  
University of Milano-Bicocca, Italy

### Reviewed by:

Maja Roch,  
University of Padua, Italy  
Vicky Chondrogianni,  
University of Edinburgh,  
United Kingdom

### \*Correspondence:

Ewa Haman  
ewa.haman@psych.uw.edu.pl  
Zofia Wodniecka  
zofia.wodniecka@uj.edu.pl

### Specialty section:

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

**Received:** 20 April 2017

**Accepted:** 09 August 2017

**Published:** 04 September 2017

### Citation:

Haman E, Wodniecka Z, Marecka M, Szewczyk J, Białecka-Pikul M, Otwinowska A, Mieszkowska K, Łuniewska M, Kołak J, Miękiś A, Kacprzak A, Banasik N and Foryś-Nogala M (2017) How Does L1 and L2 Exposure Impact L1 Performance in Bilingual Children? Evidence from Polish-English Migrants to the United Kingdom. *Front. Psychol.* 8:1444. doi: 10.3389/fpsyg.2017.01444

<sup>1</sup> Psycholinguistics Lab, Faculty of Psychology, University of Warsaw, Warsaw, Poland, <sup>2</sup> Psychology of Language and Bilingualism Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland, <sup>3</sup> Early Child Development Psychology Laboratory, Institute of Psychology, Jagiellonian University, Krakow, Poland, <sup>4</sup> Institute of English Studies, University of Warsaw, Warsaw, Poland

Most studies on bilingual language development focus on children's second language (L2). Here, we investigated first language (L1) development of Polish-English early migrant bilinguals in four domains: vocabulary, grammar, phonological processing, and discourse. We first compared Polish language skills between bilinguals and their Polish non-migrant monolingual peers, and then investigated the influence of the cumulative exposure to L1 and L2 on bilinguals' performance. We then examined whether high exposure to L1 could possibly minimize the gap between monolinguals and bilinguals. We analyzed data from 233 typically developing children (88 bilingual and 145 monolingual) aged 4;0 to 7;5 (years;months) on six language measures in Polish: receptive vocabulary, productive vocabulary, receptive grammar, productive grammar (sentence repetition), phonological processing (non-word repetition), and discourse abilities (narration). Information about language exposure was obtained via parental questionnaires. For each language task, we analyzed the data from the subsample of bilinguals who had completed all the tasks in question and from monolinguals matched one-on-one to the bilingual group on age, SES (measured by years of mother's education), gender, non-verbal IQ, and short-term memory. The bilingual children scored lower than monolinguals in all language domains, except discourse. The group differences were more pronounced on the productive tasks (vocabulary, grammar, and phonological processing) and moderate on the receptive tasks (vocabulary and grammar). L1 exposure correlated positively with the vocabulary size and phonological processing. Grammar scores were not related to the levels of L1 exposure, but were predicted by general cognitive abilities. L2 exposure negatively influenced productive grammar in L1, suggesting possible L2 transfer effects on L1 grammatical performance. Children's narrative skills benefitted from exposure to two languages: both L1 and

L2 exposure influenced story structure scores in L1. Importantly, we did not find any evidence (in any of the tasks in which the gap was present) that the performance gap between monolinguals and bilinguals could be fully closed with high amounts of L1 input.

**Keywords:** bilingual children, L1 acquisition, migrant children, Polish-English bilinguals, home language, minority language, language exposure, language input

## INTRODUCTION

Many studies examining early bilingualism in migrant populations focus on the development of the majority language<sup>1</sup> (i.e., L2, e.g., Gutiérrez-Clellen et al., 2008; Paradis, 2009; Chondrogianni and Marinis, 2011, 2012; Verhoeven et al., 2011; Hoff et al., 2012). This is because proficiency in the majority language is a prerequisite of success in education (e.g., Strand et al., 2015) and on the job market in the new country (e.g., Shields and Price, 2004; Guven and Islam, 2015). One exception to the predominance of studies on L2 is research on heritage language speakers, conducted mostly in the North American context (e.g., Montrul, 2008; Rothman, 2009; Montrul and Ionin, 2010). A heritage language is understood as “a language spoken at home or otherwise readily available to young children,” but not dominant in the larger society (Rothman, 2009, pp. 156), i.e., it is defined in the same way as we define a minority language in the current paper. While there are many studies on grammatical performance of heritage speakers in L1 (e.g., Polinsky, 2008; Rothman, 2009), there are only a few studies on other aspects of heritage language such as vocabulary and phonology (e.g., Montrul, 2010). Moreover, few of the heritage speaker studies concentrated on the language acquisition process in children (e.g., Montrul, 2008; Polinsky, 2011), but rather on its outcomes in adulthood (for discussion see Rothman, 2009; Rothman and Treffers-Daller, 2014). Overall, although research shows that maintaining the minority language (L1) is of great importance for both well-being of an individual (Portes and Hao, 1998; Yu, 2013; De Houwer, 2015) and for language preservation at the community level (Potowski, 2013), only a few studies have thoroughly examined the development and maintenance of children’s L1 (Rodríguez et al., 1995; Winsler et al., 1999; Gathercole and Thomas, 2009).

We aim to fill this gap by investigating L1 developmental patterns in migrant children raised bilingually. We focus on 4–7 year old Polish-English migrant children living in the United Kingdom. The choice of this particular language group was driven by an unprecedented influx of Poles to the United Kingdom since Poland joined the EU in 2004. The Polish community in the United Kingdom has now reached one million (White, 2011; Kulakowska, 2013), and each year c.a. 25,000 children are born to Polish families (Office for National Statistics [ONS], 2014). This offers an opportunity for systematic

and large-scale research on bilingual language development in Polish children, a community that, to our knowledge, has not been thoroughly addressed in the existing research. Although migrant communities of similar sizes exist in other countries, this one seemed especially appropriate for the purpose of studying bilingual language development because of the characteristics of this wave of Polish migration to the United Kingdom. The group, unlike most migrant communities studied so far, does not comprise exclusively unskilled workers with low socio-economic status (SES), which might bias the result. A majority of post-accession migrants from Poland to the United Kingdom were people with secondary education, many of them also holding academic degrees. Also, they were mostly young adults, often bringing young children with them or having children while staying in the United Kingdom (Okólski and Salt, 2014). For this reason, in the current paper, we take a “snapshot” of this new bilingual population and compare the bilinguals’ home language performance to that of their Polish-speaking monolingual peers raised in Poland. We also looked at the age-related differences in the two groups. We aimed to establish to what extent the bilingual migrant children and their monolingual peers in the home country differ in their L1 abilities across four domains of language, i.e., vocabulary, grammar, phonological processing, and discourse. Moreover, our goal was to examine how language experience (in both L1 and L2) influences L1 performance of bilinguals, while controlling the sources of variance related to their general cognitive abilities and socioeconomic status.

## Bilingual vs. Monolingual Language Development

Studies focusing on L2 development in bilinguals demonstrate that bilingual children lag behind their monolingual peers in most aspects of language processing, often scoring similarly to monolinguals with specific language impairment (SLI; Kohnert et al., 2009; Ebert and Kohnert, 2016). Studies investigating L1 in bilinguals offer less conclusive results (e.g., Umbel and Oller, 1994; Winsler et al., 1999), but many indicate a performance gap between bilinguals and their monolingual peers (e.g., Fabiano-Smith and Barlow, 2010). Indeed, research on adult heritage speakers indicates that literacy and formal education in the majority language (L2) often results in the incomplete heritage (L1) language acquisition (Montrul, 2008). As Sorace (2005) points out, this is because the language input heritage speakers receive varies in terms of quality, as heritage speakers are exposed to the input in the minority language mostly from their parents, whose language may have already attrited. However, the differences between monolingual and bilingual children should not conceal similarities between the two developmental paths.

<sup>1</sup> Throughout the paper, we use the term, “majority language” or L2 for the language of the country where the bilingual children of the migrants live (in the case of this study – English). We use the term “minority language,” “home language” or “L1” for the language used by at least one the child’s parents (in the case of this study – Polish).

Several studies suggest that bilinguals achieve the developmental milestones (defined as the age when the child begins acquiring a particular language skill) roughly at the same time as monolinguals do. This is true for lexical development (Pearson et al., 1993; Hoff et al., 2012), grammatical development (Paradis and Genesee, 1996; De Houwer, 2005; Genesee and Nicoladis, 2007; Paradis, 2009) and phonological development (Fabiano-Smith and Barlow, 2010). For example, both bilinguals and monolinguals utter their first words around the age of one, and have similarly sized vocabulary and phonological inventory, when both languages of the bilingual are taken into consideration (Fabiano-Smith and Barlow, 2010; Hoff et al., 2012). There is also evidence that the abilities to produce coherent discourse do not differ between bilinguals and monolinguals of a comparable age (e.g., Paradis and Kirova, 2014). In other words, there are both similarities and differences between monolingual and bilingual developmental paths. Research findings suggest that the bilingual development has its own specificity, and that monolingual norms should not be applied to bilingual speakers (Gathercole, 2013a,b; Armon-Lotem et al., 2015).

In the subsequent sections, we briefly review the literature related to the bilingual development in the four language domains that are the focus of our study: vocabulary, grammar, phonology, and discourse. For each language domain, we address two critical issues: the differences between bilingual children and their monolingual peers, and the impact of language exposure on performance in each of the language domains in L1 and L2.

## Vocabulary

Studies examining L2 vocabulary in bilingual children consistently report that bilinguals lag behind their monolingual peers on both receptive tasks (Bialystok et al., 2010; Verhoeven et al., 2011) and productive tasks (Uccelli and Pérez, 2007). Some studies even find typically developing bilingual children to have smaller receptive vocabularies in L2 than monolinguals with SLI (Verhoeven et al., 2011).

In terms of L1 vocabulary size, some studies suggest that bilingual children raised in the migrant setting are disadvantaged (e.g., Pearson et al., 1997; Uccelli and Pérez, 2007). Other studies indicate that L1 vocabulary in bilinguals is not affected negatively, either in the receptive tasks (Umbel and Oller, 1994; Winsler et al., 1999; Leseman, 2010), or in the productive tasks (Leseman, 2010). Thus, the results are inconclusive and they should be treated with caution, since the majority of L1 vocabulary studies compared children's lexical acquisition between the two languages of bilinguals, and did not compare bilinguals' L1 scores to the vocabulary scores of a matched monolingual group.

The observed discrepancy in the results on L1 vocabulary in bilinguals may stem from methodological issues (e.g., the lack of well-matched control groups), but also from the variability in exposure to languages. Previous research indicates that L1 vocabulary size is closely connected to the reported amount of L1 exposure, while L2 vocabulary size is related to exposure to L2 (Pearson et al., 1997; Vermeer, 2001; Patterson, 2002; De Houwer, 2007; Thordardottir, 2011; Hoff et al., 2012; Hoff and Core, 2013). This relationship is especially strong for the productive vocabulary. For example, in a study on English-French bilingual

children in Canada, the participants with equal amounts of exposure to L1 and L2 had native-like scores in a receptive vocabulary task, but not in a productive vocabulary task. To perform on par with the monolinguals in the productive vocabulary task, the children needed to have more exposure in the language tested (Thordardottir, 2011). Moreover, Pearson et al. (1997) established the 20% threshold hypothesis – they claim that children who hear less than 20% of their input in a given language are often unwilling to speak that language. In line with this hypothesis, Hoff et al. (2012) suggests that 20% is an absolute minimum of input for a child to be able to use a language. Studies on heritage speakers also suggest that vocabulary in L1 is affected by both the amount and quality of input in L1 (Schwartz, 2008).

Overall, the current literature indicates that bilingual children have significantly lower vocabulary scores in L2, compared to their monolingual peers, while the findings regarding L1 vocabulary are inconclusive. In general, the amount of exposure seems to be crucially linked to vocabulary performance of the bilingual children, especially in language production.

## Grammar

The studies examining specific areas of grammar in bilingual development show mixed results. On the one hand, some reported that bilinguals acquire certain structures in L2 (e.g., such as finite verb forms, Paradis and Genesee, 1996) just like their monolingual peers, especially when L2 is their dominant language (see De Houwer, 2005; Conboy and Thal, 2006; Genesee and Nicoladis, 2007; Parra et al., 2011). Still, many studies suggested that bilingual children perform worse than monolinguals on L2 grammar tasks, for example the ones examining the application of tense morphology (see Hoff et al., 2012). The bilingual disadvantage seems to be smaller for the receptive than productive tasks (Verhoeven et al., 2011; Chondrogianni and Marinis, 2012). Moreover, the majority of grammatical errors reported in studies on bilingual acquisition appear to be developmental errors (for review see Paradis, 2009). As for global L2 grammar measures, including the Sentence Repetition task (SRep; see Marinis and Armon-Lotem, 2015), which involves verbatim repetitions of sentences with various grammatical structures in the target language, bilingual children usually score lower than monolinguals (Verhoeven et al., 2011; Komeili and Marshall, 2013; Thordardottir and Brandeker, 2013). When it comes to grammatical systems of L1 in the minority speakers, they are often simplified as regards the development of certain grammatical structures (see Benmamoun et al., 2013; Scontras et al., 2015). Bilingual children can also score lower on L1 holistic grammatical assessment tasks such as the SRep, especially if they did not have much exposure to that language (Thordardottir and Brandeker, 2013). The areas of L1 grammar that appear to be particularly problematic include agreement morphology (e.g., Bolonyai, 2007; Montrul and Potowski, 2007; Polinsky, 2008; Gathercole and Thomas, 2009), overusing rigid word order patterns (e.g., Isurin and Ivanova-Sullivan, 2008), or applying and interpreting long-distance binding (e.g., Polinsky, 2006; Kim et al., 2009). However, since many of these accounts come from studies on older participants than preschool children, it is necessary to further

investigate at which point in development those alternations in syntax emerge (Polinsky, 2016). In a study focused specifically on child minority language, Montrul and Potowski (2007) investigated the acquisition of Spanish gender agreement in school-aged heritage speakers of Spanish enrolled in a dual Spanish-English immersion program. As evidenced by the data coming from an oral narrative task and a picture matching task, the heritage speakers scored lower than Spanish monolinguals but higher than the L2 learners in applying gender agreement rules to determiners and adjectives.

Overall, research indicates that poorer performance on L1 grammatical tasks might be related to impoverished or altered exposure to L1 or to the influence of the dominant community language (see Rothman, 2007; Gathercole and Thomas, 2009; Benmamoun et al., 2013; Scontras et al., 2015; Hoff et al., 2017). For example, in Spanish-English bilingual children, L1 exposure at home has been found to be related to scores in L1 (Spanish) grammaticality judgment task targeting the knowledge of gender marking and *that*-trace structures (Gathercole, 2002a,b). In Welsh-English bilinguals, home and school exposure to the L1 minority language (Welsh) correlated with children's receptive command of the syntactic patterns of Welsh gender marking and the use of word order cues in identifying subjects (Gathercole and Thomas, 2009). Montrul and Potowski (2007) observed that sequential bilinguals, who were first exposed exclusively to Spanish as an L1, and thus received more overall exposure in that language, outperformed simultaneous bilinguals in applying gender agreement rules to adjectives. The results showed that the development of certain aspects of L2 grammar may be affected by reduced exposure in early childhood.

There is also evidence suggesting that structures from the dominant language might be incorporated into the weaker language more often than the other way around (Döpke, 1998; Yip and Matthews, 2000). For instance, the effects of L2 exposure on the L1 minority language may affect some specific areas of L1 grammar, such as the use of overt versus null subjects (e.g., Paradis and Navarro, 2003), determiners (e.g. Kupisch, 2007; Montrul and Ionin, 2010) or inflectional morphology (see Benmamoun et al., 2013). However, it is often hard to disentangle the effects of L2 transfer from the effects of the reduced input in L1 (Scontras et al., 2015).

Overall, many studies suggest that bilingual children may experience developmental difficulties in the domain of morphosyntax in their non-dominant language, whether L1 or L2. Crucially, however, the gap between the performance of mono- and bilingual groups has been found to depend on the amount and type of exposure to the target language.

## Phonology

Bilingual children can differ from their monolingual peers in terms of phonological development in L1 and L2 in three ways: delay, acceleration, and transfer. First, bilinguals might learn to produce some speech patterns (e.g., vowels, Kehoe, 2002; consonants, Goldstein and Washington, 2001; prosody, Lleó, 2002) later than monolinguals. Moreover, when tested in L2 on generalized phonological assessment measures such as English

Diagnostic Evaluation of Articulation and Phonology (Dodd et al., 2002), bilingual children might obtain low scores, which in monolinguals would be typical for phonological delay (En et al., 2014). The delay in the acquisition of phonological features of L1 has also been reported (Goldstein and Washington, 2001), but not in all studies (Kehoe, 2002).

Secondly, bilinguals might acquire some phonological features in L2 faster than their monolingual peers. For instance, Polish-English bilinguals and Welsh-English bilinguals acquire complex consonantal clusters in English faster than their monolingual peers, most likely due to the fact that their L1 is rich in complex consonant clusters (Mayr et al., 2014; Tamburelli et al., 2015). To our knowledge, there have been no studies showing a similar effect for L1 in bilingual speech.

Thirdly, bilinguals might exhibit phonological transfer, i.e., pronounce the sounds in one language with the phonetic features of their other language. Phonological transfer between bilinguals' two languages may affect both prosodic patterns (Paradis, 2001) and segmental features (Fabiano-Smith and Barlow, 2010; Barlow, 2014) and can take both directions, i.e., from L1 to L2 and from L2 to L1 (Fabiano-Smith and Barlow, 2010; Fabiano-Smith and Goldstein, 2010; Marecka et al., 2016). Overall, while bilingual children do not have smaller phonological inventories than monolinguals, they tend to mix the phonological features of both languages (Fabiano-Smith and Barlow, 2010). Heritage language studies suggest that these tendencies might carry into adulthood of the bilingual speakers. L1 phonological features in the speech of adult heritage speakers such as vowel quality or VOT can shift toward L2-like values (Godson, 2004; Nagy and Kochetov, 2013), even though the L1 accent of these adult heritage speakers is reported to be more native-like than the accent of L2 learners of a particular language (Au et al., 2002; Oh et al., 2003).

Apart from testing for the ability to produce appropriate phonemes in the target language, several studies used the non-word repetition (NWR) task to study phonological processing in bilingual children. When the non-words used in the test are highly L1- or L2-like, they tend to measure the inventory of phonological representations of a child (Jones et al., 2010; Jones, 2011). Bilinguals perform worse than monolinguals on the NWR with L2-like non-words (Kohnert et al., 2006), sometimes even on par with monolinguals with SLI (Windsor et al., 2010). When tested in their L1 (and not L2) bilinguals tend to perform better (Gutiérrez-Clellen and Simon-Cereijido, 2010; Summers et al., 2010). When non-words are quasi language-universal, bilinguals perform similarly to their monolingual peers (Boerma et al., 2015).

Both phonological development and processing are influenced by the cumulative language exposure. Many studies of phonological development have reported that children who started acquiring L2 earlier (i.e., cumulatively had more exposure to L2) sound more native-like than children who started acquiring the language later (Asher and García, 1969; Snow and Hoefnagel-Höhl, 1977; Flege and Fletcher, 1992; Flege, 1995; Aoyama et al., 2008). Moreover, the phonological performance in both L2 and L1 is directly proportional to the exposure and use of a particular language (Flege, 2002). Phonological processing

(as measured with NWR) is also connected to the amount of exposure that bilinguals receive in the tested language (Summers et al., 2010), although to a smaller degree than vocabulary (Thordardottir and Brandeker, 2013).

## Discourse

In studies of discursive abilities, children are usually asked to narrate a story, often based on pictorial stimuli. Narrative data support the results from standardized tests by providing additional performance measures across the languages of the bilingual child (Iluz-Cohen and Walters, 2012). A measure usually taken into consideration here is the structural coherence of narratives, i.e., the story structure, which is subsequently assessed in terms of how well the child refers to the goals of the characters, the attempts to reach these goals and their outcomes (Gagarina et al., 2016; see also Stein and Glenn, 1979). Story structure scores go beyond the assessment of single words or sentences, but instead indicate the level of more complex cognitive and pragmatic abilities (Gagarina, 2016). Studies comparing the story structure of bilinguals in L2 or L1 with that of their monolingual peers are infrequent and their results are mixed. One study comparing L1 structural coherence in bilingual Finnish-Swedish children with that of Finnish monolinguals found no differences between the two groups of children (Kunnari et al., 2016). On the other hand, in a study comparing the performance in L1 Russian of Russian-Norwegian children to Russian monolinguals, the bilinguals scored lower on the story structure in their L1 (Rodina, 2016). The same pattern has also been observed in the studies on heritage speakers. In a case study by Polinsky (2008), two heritage speakers of Russian (a 9-year-old and a college student) were found to produce significantly shorter utterances and narrate at a slower pace than monolingual Russian speakers.

The effect of language exposure on children's narrative abilities is a complex issue. On the one hand, some findings suggest that the exposure to a particular language might not be crucial to narrating in that language. Most studies comparing bilingual children's narrative abilities in L1 and L2 indicate that the structure of narratives is relatively invariant across languages and that the measures of the story coherence in the child's two languages tend to be highly correlated (Muñoz et al., 2003; Fiestas and Peña, 2004; Uccelli and Pérez, 2007; Gagarina, 2016; Kunnari et al., 2016). In general, children produce equally coherent stories in both languages, even if the child's linguistic abilities in terms of vocabulary or grammar in one of the languages are weaker (Gagarina, 2016). The finding that the story structure does not differ across the languages of a bilingual is probably related to the fact that the ability to tell coherent stories taps into the child's general knowledge about the world and thus seems to be relatively language-independent (Gagarina, 2016; Gagarina et al., 2016). This would indicate that language-specific exposure might not be crucial for developing narrative skills.

On the other hand, several studies point to the importance of language exposure, showing that the narrative structure in bilinguals might be better in L1 than in L2 (Kapalková et al., 2016; Roch et al., 2016). A study on L1 Russian narratives in Russian-Norwegian preschoolers suggests that the L1 story

structure might be dependent on the amount of exposure to L1 (e.g., Rodina, 2016). Further, as indicated by Gagarina (2016), the strong positive correlations between the story structure in L1 and L2 cease to occur after several years of schooling in the majority language. Then, the stories told in the language of schooling become more coherent than those in the home language. This result suggests that the story structure, rather invariant across languages in young bilinguals, might be sensitive to explicit narrative teaching at school and to receiving large amounts of structured input and modeling in the majority language. Finally, several studies showed that older bilingual children produce more coherent stories than younger children (Bohnacker, 2016; Maviş et al., 2016). This might be attributable to children's cognitive maturity, but also to the differences in language exposure.

To conclude, bilingual children's discursive abilities are rather under-researched in comparison with other aspects of language use, and the results of studies are not clear-cut. Some suggest that the narrative abilities of bilinguals might be influenced by exposure and modeling, especially at the later stages of education. However, the results of studies on the narrative abilities in bilingual preschool children suggest that producing coherent stories is an area where bilinguals and monolinguals might perform similarly, regardless of the L1 exposure.

## The Current Study

The literature review presented above reveals a rich body of research devoted to language acquisition in bilingual children. However, it is clear that despite the wealth of studies, many facets of bilingual language acquisition are still under-researched. The majority of studies focused on the L2 of bilinguals and only few examined their L1 and benchmarked it against a monolingual control group (e.g., Umbel and Oller, 1994; Thordardottir, 2011; Thordardottir and Brandeker, 2013). Moreover, only few studies investigated several different language measures on the same group of participants (Uccelli and Pérez, 2007; Verhoeven et al., 2011; Thordardottir and Brandeker, 2013). Thus, there is certainly a need for large-scale investigations that would allow to obtain a comprehensive picture of differences in the linguistic performance between monolinguals and bilinguals by comparing them in different areas of language use. Also, a certain limitation of many previous studies is that they seldom controlled for language exposure in the bilingual group, despite the fact that this single variable can potentially explain many differences between monolinguals and bilinguals (Pearson et al., 1997; Thordardottir, 2011; Thordardottir and Brandeker, 2013). Finally, to our best knowledge, there are no studies which would examine the effect of language exposure on different language domains in child bilingual speakers, while controlling for potentially confounding variables such as short-term memory (STM) capacity, non-verbal IQ, or SES. Controlling these variables seems important, since research consistently indicates their crucial role in language development. STM capacity has been linked to the development of vocabulary (Gathercole et al., 1992) and both vocabulary and grammar (Verhagen and Leseman, 2016) in preschool children. Moreover, deficits in non-verbal IQ might be linked to language deficits (Botting, 2005) and SES might determine the overall language development (see Hoff, 2006 for a review; Hoff, 2013).

## Measuring Language Exposure

Although it is generally agreed that language exposure plays an important role in language acquisition, the construct is a matter of much controversy (Carroll, 2017). The term “language exposure” lacks an accurate definition and is measured in various ways (see Armon-Lotem, 2016; Carroll, 2017 for discussion). In the present paper, we are following Carroll (2017) and we define exposure as an observable and measurable contact with a particular language.

The quantification of language exposure has been a challenging task. To estimate exposure several related factors can be used: the intensity of contact with a given language (also as a function of the number of interlocutors available for a given language), the age of the first contact with the language, and the time spent while exposed to a particular language. Indirectly, also chronological age might be a contributing factor, because older children tend to have greater length of exposure to a given language in their lifetime. Ideally, all these factors should be disentangled and their contribution measured independently. However, because these predictors are highly correlated, doing so would require testing huge participant samples, and to the best of our knowledge, no study has accomplished this so far. The existing studies that controlled for one of these factors conceded that the other ones were left uncontrolled (e.g., Bedore et al., 2016). One way of solving this problem is to eliminate at least one factor, for example the Age of Acquisition, by testing populations that are exposed to both languages from birth (e.g., testing English-French in bilingual families in Montreal; Thordardottir, 2017). But even then, the contribution of the three other highly correlated variables remains to be controlled. A better way of addressing the problem is to circumvent it by creating one cumulative index that encompasses all the related factors. Such an approach was taken in a few recent studies (Unsworth, 2013; Unsworth et al., 2014; Vender et al., 2016) and it is also chosen in the present study. Such an index typically reflects the length of exposure to a language (from the age of the first contact to the time of testing), obtained from parental questionnaires. Specific approaches to exposure may differ in how exactly this information is elicited via background questionnaires. For example, Unsworth (2013) estimates the percentage of waking hours during which children were exposed to a particular language, in each year of their life. In the present study, we estimated the intensity of contact with Polish and English. We multiplied this estimation by the time before and after migration, respectively. The estimate of intensity of contact was based on the number of speakers at home when the language was used. Hence, our index of language exposure simultaneously reflects both the quantity and quality of exposure (i.e., the number of different speakers). In the methods section, we describe how our index of cumulative language exposure was constructed in more detail.

## Research Questions

Here, we present a comprehensive analysis of L1 performance in bilingual migrant children, as compared with their monolingual peers, with a number of factors controlled. We used six direct language measures to test over 200 typically developing children

(including more than 80 bilinguals) aged 4;0 to 7;5. The measures included receptive and productive vocabulary, receptive and productive grammar (SRep), phonological processing (NWR), and narrative skills. What is more, in the current analyses, we assess the impact of exposure to both L1 and L2 on bilinguals' performance in each of the language domains.

Our analyses focused on the three main research questions:

- (1) What are the differences between bilingual migrant children and their monolingual peers in the four domains of Polish L1 development?
- (2) How does the cumulative exposure to L1 and the cumulative exposure to L2 influence performance of the bilingual children in each of the language domains?
- (3) Can high exposure to L1 minimize the potential gap between monolinguals and bilinguals?

## MATERIALS AND METHODS

### Participants

Overall, 173 bilingual children and 311 monolingual participants took part in the study. However, the analyses presented in the current paper were based on subsamples from both groups. In the analyses, we considered only those participants for whom we had a full data set necessary to control for the non-verbal intelligence (*Raven's Colored Progressive Matrices*; Raven, 2003; Jaworowska and Szustrowa, 2003), STM (forward digit span, Wechsler, 1974), and SES (background questionnaires). We excluded the children who had hearing problems (6 bilinguals, 3.5% of the bilingual sample; 9 monolinguals; 2.9% of the monolingual sample). Additionally, from the bilingual group we excluded the children who were effectively trilingual (15 children; 8.7% of the bilingual sample; see also Mieszkowska et al., 2017), from the monolingual group those who occurred to be bilingual (living in Poland, 3 children, 1% of the monolingual sample) and those at risk of SLI, as indicated by parental concerns reported in the questionnaires (4 bilinguals; 2.3% of the bilingual sample, 3 monolinguals, 1% of the monolingual sample). Eventually, data from 233 children (88 bilingual and 145 monolingual) were considered for further analyses. Seventy of the bilingual children who took part in the study had both Polish-speaking parents. Eighteen children lived in families with a Polish-speaking mother and a father speaking English at home (11 native English speakers and 7 non-native English speakers). All the bilinguals lived in the United Kingdom, but they varied in terms of the age of their first contact with English ( $M = 13$  months,  $SD = 16$  months). Fifty-five of them were first exposed to English within the first year of life (36 just after birth). Others had their first contact with English later (up to 60th month of life).

For each of the language measures reported in this paper, we conducted separate analyses on a subsample of children. The subsamples consisted of all bilingual children for whom we had the data on the task of interest and a group of monolinguals matched one-to-one to the bilingual group on age, SES (years of mother's education), gender, non-verbal IQ (Raven scores), and STM (as measured by forward digit span).

TABLE 1 | Demographic information and descriptive statistics for background measures in the participant subsamples.

|                                                           | In total              |       | Subsample 1<br>(Receptive<br>vocabulary) |       | Subsample 2<br>(Productive<br>vocabulary) |       | Subsample 3<br>(Receptive<br>grammar) |       | Subsample 4<br>(Productive<br>grammar) |       | Subsample 5<br>(Phonological<br>processing) |       | Subsample 6<br>(Discourse) |       |
|-----------------------------------------------------------|-----------------------|-------|------------------------------------------|-------|-------------------------------------------|-------|---------------------------------------|-------|----------------------------------------|-------|---------------------------------------------|-------|----------------------------|-------|
|                                                           | Mean                  | SD    | Mean                                     | SD    | Mean                                      | SD    | Mean                                  | SD    | Mean                                   | SD    | Mean                                        | SD    | Mean                       | SD    |
| <b>Bilinguals</b>                                         | <b>N = 88 (51 F)</b>  |       | <b>N = 87 (50 F)</b>                     |       | <b>N = 87 (50 F)</b>                      |       | <b>N = 74 (42 F)</b>                  |       | <b>N = 80 (46 F)</b>                   |       | <b>N = 79 (47 F)</b>                        |       | <b>N = 53 (32 F)</b>       |       |
| Age (years; range: 4.36–7.01)                             | 5.69                  | 0.76  | 5.69                                     | 0.76  | 5.69                                      | 0.76  | 5.76                                  | 0.74  | 5.77                                   | 0.73  | 5.76                                        | 0.75  | 5.71                       | 0.71  |
| Years of mother's education<br>(range: 10–24)             | 16.27                 | 2.98  | 16.32                                    | 2.96  | 16.32                                     | 2.96  | 16.30                                 | 3.15  | 16.45                                  | 2.93  | 16.30                                       | 2.89  | 16.40                      | 2.81  |
| Raven (range: 12–34)                                      | 21.91                 | 5.83  | 21.89                                    | 5.86  | 21.89                                     | 5.86  | 21.86                                 | 5.63  | 22.31                                  | 5.83  | 22.25                                       | 5.73  | 22.02                      | 5.10  |
| Digit span (range: 0–6)                                   | 3.90                  | 1.12  | 3.91                                     | 1.13  | 3.91                                      | 1.13  | 3.88                                  | 1.15  | 3.98                                   | 1.04  | 4.01                                        | 0.97  | 4.08                       | 0.83  |
| Age of first L2 contact (years;<br>range: 0–5.0)          | 1.08                  | 1.34  | 1.10                                     | 1.34  | 1.10                                      | 1.34  | 1.17                                  | 1.40  | 1.17                                   | 1.37  | 1.16                                        | 1.38  | 1.09                       | 1.32  |
| L1 (Polish) cumulative exposure<br>(range: 70.83–515.86)  | 316.45                | 93.64 | 316.39                                   | 94.19 | 316.39                                    | 94.19 | 319.93                                | 99.05 | 326.11                                 | 89.54 | 323.68                                      | 91.91 | 317.37                     | 84.66 |
| L2 (English) cumulative exposure<br>(range: 16.87–362.13) | 158.85                | 81.34 | 158.36                                   | 81.68 | 158.36                                    | 81.68 | 158.74                                | 86.11 | 155.11                                 | 82.35 | 157.44                                      | 82.39 | 155.88                     | 77.68 |
| <b>Monolinguals</b>                                       | <b>N = 145 (74 F)</b> |       | <b>N = 87 (50 F)</b>                     |       | <b>N = 87 (50 F)</b>                      |       | <b>N = 74 (42 F)</b>                  |       | <b>N = 80 (45 F)</b>                   |       | <b>N = 79 (47 F)</b>                        |       | <b>N = 53 (31 F)</b>       |       |
| Age (years; range: 3.52–7.23)                             | 5.60                  | 0.72  | 5.64                                     | 0.70  | 5.63                                      | 0.69  | 5.67                                  | 0.74  | 5.60                                   | 0.69  | 5.70                                        | 0.63  | 5.73                       | 0.66  |
| Years of mother's education<br>(range: 11.5–25)           | 17.46                 | 2.69  | 16.74                                    | 2.43  | 16.80                                     | 2.33  | 16.66                                 | 2.58  | 16.81                                  | 2.77  | 16.88                                       | 2.44  | 16.91                      | 2.96  |
| Raven (range: 11–32)                                      | 22.05                 | 5.00  | 21.90                                    | 4.69  | 21.91                                     | 4.84  | 22.36                                 | 4.60  | 22.23                                  | 4.29  | 22.30                                       | 4.63  | 22.40                      | 5.39  |
| Digit span (range: 0–7)                                   | 3.95                  | 0.93  | 3.97                                     | 0.99  | 3.98                                      | 1.02  | 4.00                                  | 1.01  | 4.01                                   | 0.92  | 4.08                                        | 0.94  | 4.15                       | 0.97  |

The matching procedure served to ensure that any differences between the groups can be attributed to language status (bilingual or monolingual), and not to other factors known to affect the performance in the tasks of interest, such as environmental differences related to SES (see Hoff, 2006; Qi et al., 2006; Hoff and Core, 2013), or children's cognitive abilities (see Kail, 2000). The characteristics of the overall sample and the task-specific subsamples are presented in **Table 1**.

## Materials and Procedures

### Tasks

The testing battery included six published normed tests or their non-normed adaptations, six experimental tasks used in previous research, six language tasks designed as a part of the Bi-SLI-Poland project within the European COST Action IS0804, and three experimental tasks designed for the project. Below all the tasks are recounted and the tasks used in the current analysis which do not have standardized administration procedures described in the tests manuals are presented in more detail.

#### *Receptive vocabulary (Obrazkowy Test Słownikowy – Rozumienie, OTSR)*

Children's receptive vocabulary was measured with *Obrazkowy Test Słownikowy*, OTSR (The Picture Vocabulary Test – Comprehension; Haman and Fronczyk, 2012). Each child was tested with two available versions of the test (A and B) to allow more data points in the assessment. The two versions of the test are fully comparable with each other and are used independently when testing for diagnostic purposes or when a retest is needed in a short period of time. Each version includes 88 items that are ordered from the least to the most difficult. The OTSR assesses the comprehension of nouns, verbs, and adjectives. Each test item is accompanied by four colored pictures. One picture depicts the target word and the three other pictures are foils, which consistently include one phonetic foil, one semantic foil, and one thematic foil.

The child is presented with one word at a time and has to point to one picture out of four that appropriately depicts the word. The child does both versions of the test, with the order of the versions counterbalanced. Depending on the child's age, the easier, initial items are skipped in each version. The procedure in each version is terminated after four consecutive errors.

Overall, a participant can receive a maximum of 88 points in each version – one point for each correct answer. For the purpose of this study, we considered only one of the test versions, for which a child obtained a higher score. We assumed that this score was more immune to the problems connected with test delivery, such as the child's boredom, or lack of concentration that led to the early termination of the test.

#### *Productive vocabulary (Zadanie Nazywania Obrazków, ZNO)*

The productive vocabulary was measured with *Zadanie Nazywania Obrazków*, ZNO (Picture Naming Task; Haman et al., 2012; Haman and Smoczyńska, 2010, unpublished). The task consists of 53 color pictures depicting 32 nouns and 21 verbs presented in the order of ascending difficulty. Each child is

presented with all 53 pictures one by one, and is asked to name each picture with one word. The task has to be administered to the last item, regardless of the number of errors made by the child. The child scores a point for each correct answer, which includes the target word, its close synonym, or a dialectal variant. The maximal number of points is 53.

#### *Receptive grammar (TROG-2)*

We used the Test for the Reception of Grammar – TROG-2 (Bishop, 2003; the Polish translation by Smoczyńska, 2008, unpublished) as a measure of receptive grammar. TROG-2 tests the comprehension of 20 syntactic constructs, organized in blocks A–T with progressing order of difficulty, as established for the English version. Each grammatical construct is included in four test items. The structures tested by TROG-2 include, for example: negatives, singular and plural inflection, object and subject relative clauses, etc. (for the exhaustive list of TROG-2 structure blocks, see Bishop, 2003).

Each test item is presented in a multiple-choice format with four pictures presented on a single board. One of the pictures illustrates the target structure and three constitute the lexical and grammatical foils to this structure. The child is auditorily presented with the stimulus containing a particular grammatical structure. Then the experimenter asks the child to point to one of four pictures which best corresponds to what he/she has heard. For each correct answer the child scores one point, and the maximum number of points is 80. In the Polish version of TROG-2 all children were expected to complete the entire task.

#### *Productive Grammar (Sentence Repetition, LITMUS-SRep)*

Productive grammar was examined with the Polish adaptation of Sentence Repetition task, LITMUS-SRep (henceforth: SRep, Banasik, Haman, and Smoczyńska, 2012, unpublished), based on the English task SASIT (Marinis et al., 2010). The adaptation is composed of 68 Polish sentences, with varying levels of grammatical complexity. The sentences contain a wide range of grammatical constructions, including negations, questions, passives, object and subject relative clauses, conditionals, object and subject clefts and noun complement clauses. The sentences are morphologically varied and controlled for length (between 5 and 9 words, no more than two clauses) and the properties of the content words used (lexical frequency, age of acquisition). All the sentences were recorded by two native speakers of Polish (male and female).

During task administration, children are asked to listen to the recorded sentences one by one and repeat them as accurately as possible. Each sentence is heard only once. The child is praised for repeating the sentences irrespective of accuracy, but no corrective feedback is given. The repetitions are recorded and then transcribed. The final score reflects the percent of correctly repeated words, relative to all the words in a given sentence (range 0–100).

#### *Phonological processing (Non-word Repetition, NWR)*

We tested phonological processing with the Polish NWR task, NWR (Szewczyk and Wodniecka, 2012), consisting of 50 non-words. All non-words, recorded by a female native speaker

of Polish, are between 2 and 4 syllables long, have a fixed stress pattern on the penultimate syllable (which is the default stress pattern in the Polish language) and are phonotactically legal. Most of the items are highly Polish-like, i.e., they contain consonant clusters and affixes typical for Polish morphology. Sometimes, they also contain lexical morphemes. The recordings of non-words are presented in the order of increasing difficulty. Participants listen to the recordings via headphones and repeat them. Subsequently, the recorded repetitions are transcribed by two independent judges. Based on their transcriptions, each non-word is categorized as either correct or incorrect. Developmental errors are disregarded and treated as correct productions. For each correctly repeated word the child receives one point. The maximal number of points for this task is 50.

### ***Discourse (LITMUS-Multilingual Assessment Instrument for Narratives, LITMUS-MAIN)***

To assess children's discursive abilities we used the Polish adaptation of the LITMUS-Multilingual Assessment Instrument for Narratives, LITMUS-MAIN (henceforth: MAIN; Gagarina et al., 2012) by Kiebzak-Mandera et al. (2012). The MAIN consists of four parallel cross-culturally neutral picture stories, each comprising six pictures. Each story includes three episodes (two pictures per episode). The episodes can be described in terms of the GAO sequences: a Goal (i.e., the protagonist wanting something), an Attempt to reach this goal, and the Outcome (e.g., The cat wants to catch a butterfly – Goal; The cat jumps forward – Attempt; The cat falls into the bushes – Outcome). The testing procedure involved two modes, the Telling mode and the Retelling mode.

Each session starts with a warm-up conversation, followed by the Telling mode and the Retelling mode. In the Telling mode, the experimenter presents the child with three envelopes, containing the same picture story. The child is asked to choose one envelope, look at the pictures and tell a story based on the pictures without showing them to the experimenter (the non-shared attention paradigm). In the Retelling mode, the experimenter shows the child another picture story, tells the story to the child and asks the child to retell the story based on the pictures and the model story he/she has heard (the shared attention paradigm). The whole session is recorded and transcribed.

In this study, we assessed the story structure of each narrative (told and retold) in accordance with the MAIN (see Gagarina et al., 2012). The child could get the maximum of 2 points for the setting of the story and then 5 points for each episode including the GAO sequences (1 point for conveying the initial mental state of the character, 1 point for expressing the Goal, 1 point for the Attempt, 1 point for the Outcome, and 1 point for describing the character's reaction to the outcome), which gives the maximum of 17 points per story.

### **Procedure**

All children were tested individually in a quiet room: the monolingual Polish children in their preschools or in their homes in Poland, the bilingual children in their schools or in their homes in the United Kingdom. Apart from the language tasks in Polish described above, each bilingual child was tested with a set of

analogous language tasks in English, but these tasks are beyond the focus of the present report. Moreover, all children were tested with a battery of cognitive tasks, including the Digit Span (Wechsler, 1974)<sup>2</sup> and Raven's Colored Matrices (Jaworowska and Szustrowa, 2003). The bilingual children were tested on the cognitive tasks only in their dominant language, as declared by their parents. In the case of children whose parents declared that they could not indicate which language was dominant, it was assumed that the child was balanced in their knowledge of the two languages and the language in which the cognitive tasks were performed was randomly selected.

Each monolingual child was tested throughout 3–4 sessions and each bilingual child – throughout 5–7 testing sessions (2–3 sessions in the non-dominant language and 3–4 sessions in the dominant language). Each session lasted approximately 45–90 min including breaks between the tasks. The duration of the session depended on the child's pace of doing the tasks. The order of the tasks in the testing sessions was counterbalanced across participants. The tasks in Polish were administered by a native speaker of Polish, while the tasks in English (not included in the present report) were administered by a native speaker or a highly proficient user of English. Polish and English were never tested on the same day.

### **Calculating the Index of Cumulative Exposure to L1 and L2**

In order to statistically control for the language exposure of bilingual children, we calculated an index of cumulative language exposure in L1 and L2. First, we estimated to what extent a child was exposed to each language when living in the United Kingdom on the basis of the Questionnaire for Parents of Bilingual Children [PABIQ – Tuller, 2015; Polish adaptation by Kuś, Otwinowska, Banasik, and Kiebzak-Mandera (2012, unpublished)]. In the questionnaire, we asked parents to estimate on a 5 point Likert scale how often the child was addressed in English and Polish in particular communicative situations such as parents talking to the child, other children talking to the child in the day-care, etc. (0 – not at all, 4 – exclusively in this language)<sup>3</sup>. These scores were aggregated to obtain an estimate of the bilingual children's exposure to Polish and to English during their stay in the United Kingdom. The maximal score for each language was 91, the actual values for L1 (Polish) were in the 15–67 range ( $M = 45.93$ ,  $SD = 11.63$ ), and for L2 (English) in the 15–61 range ( $M = 36.01$ ,  $SD = 11.31$ ). Because some of bilingual children (16 participants) in our group were born in Poland and only later immigrated to the United Kingdom, we assumed that when living in Poland the children had the maximal exposure to Polish (i.e., 91) and none to English. After immigrating to the United Kingdom, some children regularly spent a considerable amount of time in Poland (e.g., 3 months of summer holidays

<sup>2</sup>We slightly modified the original instruction to make it friendlier for children younger than 6-year-olds.

<sup>3</sup>The issues concerning the exposure to English and to Polish were not interdependent. More specifically, the parents could indicate that the child had a large exposure to both L1 and L2, or that the child had little exposure to both languages. In consequence, the estimates of exposure to L1 and L2 were only moderately correlated ( $r = -0.56$ ,  $p < 0.001$ ).

each year). Thus, we assumed that also during these periods of time the children had the maximal exposure to Polish and no exposure to English.

The final index of cumulative exposure reflected the time spent in Poland and in the United Kingdom in the lifetime of each child, as well as the amount of exposure the child received in each of these countries. The index of the cumulative exposure to Polish was calculated using the following formula: (time<sup>4</sup> spent in Poland) \* 91 + (time spent in the United Kingdom) \* (exposure to Polish while in the United Kingdom). The actual unit of measurement used to calculate the index was the child's age in days represented as years (in decimals). The mean cumulative exposure to Polish was 316.45 (*SD* = 93.64, range: 70.83–515.86). The index of cumulative exposure to English was calculated as: (the time spent in Poland) \* 0 + (the time spent in the United Kingdom) \* (the exposure to English while in the United Kingdom). The mean cumulative exposure to English was 158.85 (*SD* = 81.34, range: 16.87–362.13). **Figure 1** shows different possible scenarios of how language exposure can change with age influencing values of the cumulative exposure index.

The index of exposure will be used only in the regression analyses focusing on the bilingual group, which is the main focus of the present paper. We could not directly compare the monolingual and the bilingual groups with regards to exposure, because only parents of the bilingual children filled in the questionnaire concerning exposure to both languages.

## Statistical Analyses

As indicated earlier, in the analyses we focused on three central questions: (1) What are the differences between bilingual migrant children and their monolingual peers in the four domains of Polish L1 development? (2) How does the cumulative exposure to L1 and the cumulative exposure to L2 influence performance of the children in each language domain? (3) Can high exposure to L1 minimize the potential gap between monolinguals and

bilinguals? To address the first question, we conducted a series of independent *t*-tests to compare the average scores of the bilingual and the one-to-one matched monolingual samples. To address the remaining questions, for each task we conducted a multiple regression analysis, exclusively on the bilingual sample. For the regression analyses we used the all-subsets method with *regsubsets()* function in the *leaps* package in R (Lumley and Miller, 2004) which performs an exhaustive search for the best regression model, containing a subset of predictors used in the maximal model. The maximal model contained cumulative exposure to Polish and the cumulative exposure to English as predictors, alongside with age, years of mother's education, forward digit span, and Raven raw scores. The four latter factors were entered into the model to control for possible confound variables connected with cognitive development and SES. All the analyses were conducted on the subsamples of children to maximize the number of data points in the models – and thus the statistical power.

To test whether high exposure to Polish can minimize any performance gaps between monolinguals and bilinguals, for each task, we conducted additional analyses in which we selected a subset of 50% bilingual children with the highest weighted exposure<sup>5</sup> to L1 (or the lowest exposure to L2, if L2 exposure was the significant factor) and compared them against their monolingual peers matched one-to-one (an analysis comparing the two groups on the full set of participants was not possible, see footnote 4). This regression analysis included two variables: Age and Group (monolingual, bilingual), and the interaction of Age and Group. A significant interaction would indicate that the magnitude of the gap between the groups changes with age.

To depict the effects of exposure and to visualize the comparison of performance between the monolingual and the bilingual group, for each task we overlaid the best-fit regression lines for the two groups, as a function of age (**Figures 2–6**). For the bilingual group, the regression line is broken down by

<sup>4</sup>Our measure took into the account not only the years, but also the months and the days. The months and the days were represented in decimal values. For instance, a child could spend 2.42 years in the United Kingdom (i.e., 2 years, 5 months, and 3 days).

<sup>5</sup>The weighted estimate of exposure is simply the cumulative exposure in a given language divided by age. We use this index (rather than the cumulative exposure index) as a base of the median split for the purpose of visualization, because the graphs are plotting the data already as a function of age.

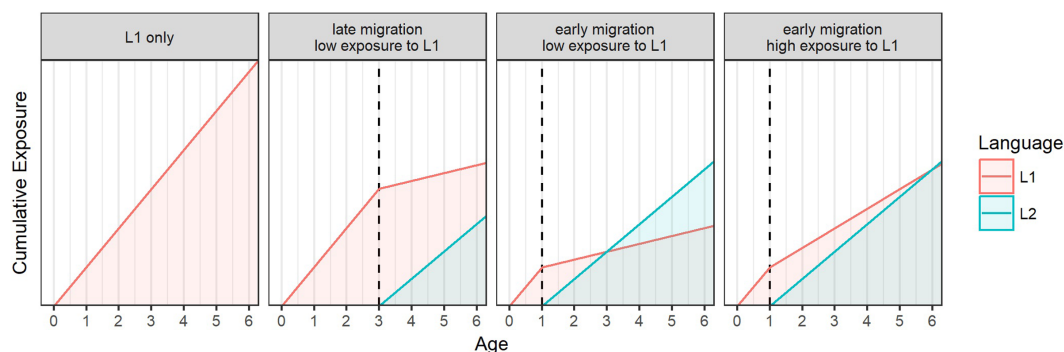
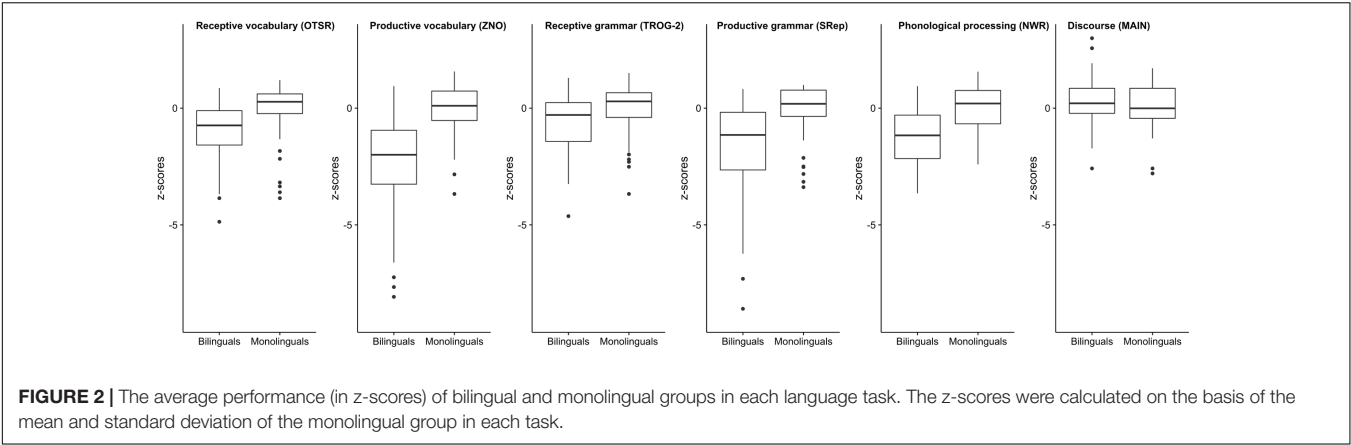


**FIGURE 1 |** Four examples of how cumulative exposure to L1 and to L2 may change with age. The intensity of exposure corresponds to the line slopes, whereas the position on the Y-axis corresponds to cumulative exposure. The examples vary with respect to the age of migration (age of contact with L2, indicated by the dashed line) and the subsequent intensity of exposure to L1.



a weighted estimate of exposure to Polish<sup>5</sup> and this is consistently done for all graphs, regardless of whether cumulative language exposure to Polish turned out to be a significant predictor in the model. Additionally, whenever cumulative language exposure to English turned out to be a significant predictor in the model, we added a graph where the regression line is broken down by a weighted estimate of exposure to English.

RESULTS

**Figure 2** presents box plots showing the average performance of bilingual and monolingual groups in each language task. Although all the analyses were conducted on the raw scores, the graphs present the results converted to z-scores to allow easier comparison across different language measures. The z-scores were calculated on the basis of the mean and standard deviation of the monolingual group in each task.

Receptive Vocabulary Test (OTSR)

On the receptive vocabulary task, the bilinguals scored on average 59.79 points out of 88 (*SD* = 14.03, range: 14–82), while the monolingual group scored on average 71.77 points (*SD* = 11.87, range: 26–86). The effect size as measured by Cohen’s *d* was large ( $t(172) = 5.99, p = 0.000, 95\% \text{ CI } [7.92, 15.69], \text{Cohen’s } d = 0.91$ ).

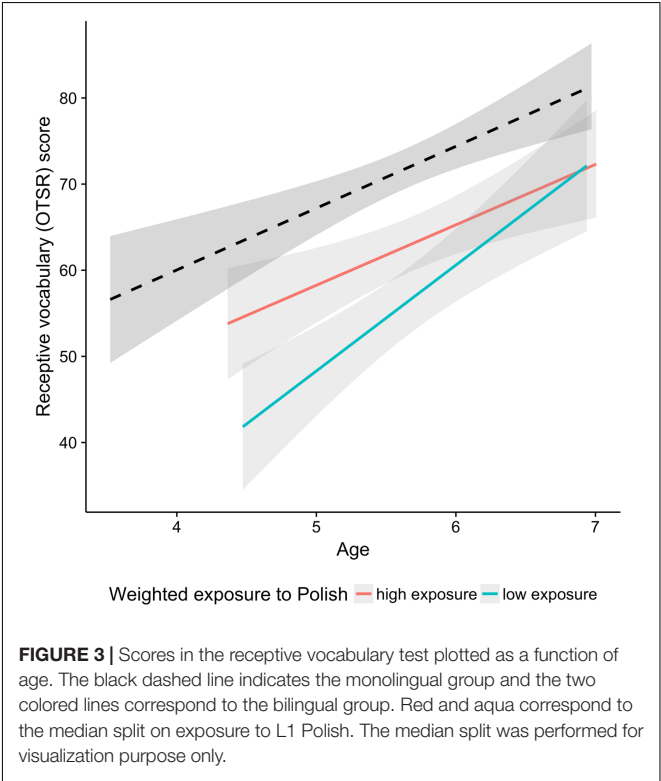
**Table 2** presents the best regression model predicting the scores on the receptive vocabulary test in the bilingual group. The significant predictors in the model were Raven, digit span, and Polish cumulative exposure: the higher score in vocabulary test was related to higher IQ score, higher digit span, and greater cumulative exposure to Polish.

**TABLE 2 |** The best regression model predicting the receptive vocabulary in the bilingual group.

|                                 | Estimate | SE   | t    | P     |
|---------------------------------|----------|------|------|-------|
| Intercept                       | 15.04    | 5.59 | 2.69 | 0.009 |
| Raven                           | 0.96     | 0.22 | 4.38 | 0.000 |
| Digit span                      | 3.20     | 1.10 | 2.91 | 0.005 |
| L1 (Polish) cumulative exposure | 0.04     | 0.01 | 2.82 | 0.006 |

$F(3,83) = 22.92. p < 0.001, \text{Adj. } R \text{ squared} = 0.43.$

**Figure 3** shows the difference in the receptive vocabulary scores depending on age, the amount of L1 exposure and group. A visual inspection of the figure suggests that a gap between the bilingual and monolingual children does not diminish with age, even in children with high exposure to Polish. A regression analysis with 50% of bilingual children with highest weighted exposure to Polish and their monolingual peers confirmed that the size of the gap between the monolingual and the high-exposure bilingual group does not diminish with age: There were significant main effects of Age and Group ( $p < 0.001$ ), but no interaction ( $p > 0.3$ ). The same type of regression analysis was repeated for other language tasks and is reported in the subsequent sections.



## Productive Vocabulary Task (ZNO)

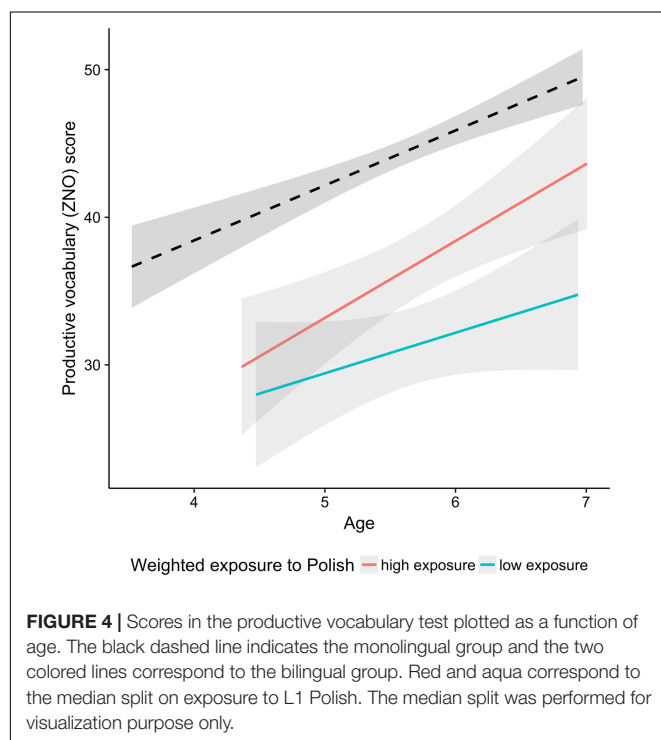
On the productive vocabulary test, the bilingual group scored on average 34.13 points out of 53 ( $SD = 8.91$ , range: 6–49), while the monolinguals scored 44.52 points ( $SD = 4.77$ , range: 27–52). The difference between the groups was statistically large ( $t(172) = 9.59$ ,  $p = 0.000$ , 95% CI [8.25, 12.53], *Cohen's d* = 1.45).

**Table 3** presents the best regression model predicting the scores on the productive vocabulary task in the bilingual group. The significant predictors in the model were the Raven's test scores and Polish cumulative exposure: the children with higher IQ, as well as those with higher cumulative exposure to Polish, had higher scores on the productive vocabulary test. **Figure 4** shows the increase in the scores with age for both monolinguals and bilinguals. Although the age-related increase in performance can be observed for children with both high and low levels of exposure to Polish, the children with high L1 exposure seem to benefit more. Still, there is a visible gap in performance between the monolinguals and bilinguals. A regression analysis with 50% of bilingual children with highest weighted exposure to Polish and their monolingual peers showed significant main effects of Age and Group ( $p < 0.001$ ), but no interaction

**TABLE 3 |** The best regression model predicting the productive vocabulary in the bilingual group.

|                                 | Estimate | SE   | t    | P     |
|---------------------------------|----------|------|------|-------|
| Intercept                       | 12.49    | 3.52 | 3.48 | 0.001 |
| Raven                           | 0.42     | 0.14 | 2.88 | 0.005 |
| L1 (Polish) cumulative exposure | 0.04     | 0.01 | 4.43 | 0.000 |

$F(2,84) = 20.20$ ,  $p < 0.001$ , Adj. *R squared* = 0.31.



( $p > 0.6$ ). Therefore, while the gap between monolinguals and bilinguals seems smaller for the bilingual group with higher levels of exposure to Polish, the additional analyses do not provide any evidence that at high levels of L1 exposure, the gap can significantly decrease at later age.

## Receptive Grammar Test (TROG-2)

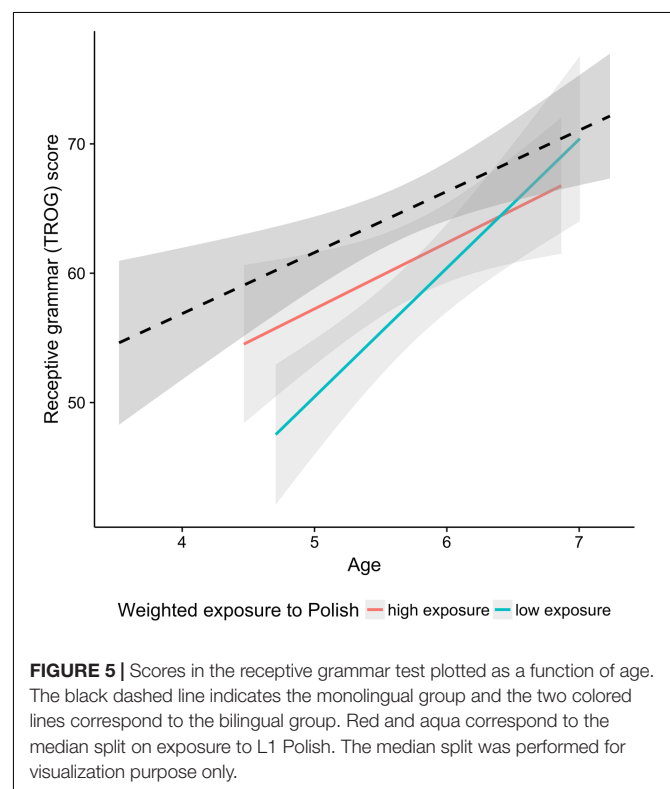
On the receptive grammar task, the bilingual group scored on average 59.46 points out of 80 ( $SD = 10.86$ , range: 21–77), while the monolinguals scored 64.76 points ( $SD = 9.46$ , range: 30–79). The difference between the two groups was significant with a medium effect size ( $t(146) = 3.16$ ,  $p = 0.002$ , 95% CI [1.99, 8.61], *Cohen's d* = 0.52).

**Table 4** shows that the TROG scores were predicted by the Raven's test scores and the digit span scores. Children who had higher scores on these tasks performed better on the receptive grammar test. Cumulative exposure to L1 (Polish) or L2 (English) was not included in the final model. As indicated by **Figure 5**, the gap in scores between the monolingual children and bilinguals is not very large and seems to decrease with age, particularly for the

**TABLE 4 |** The best regression model predicting the receptive grammar in the bilingual group.

|            | Estimate | SE   | t    | P     |
|------------|----------|------|------|-------|
| Intercept  | 27.47    | 4.36 | 6.30 | 0.000 |
| Raven      | 0.83     | 0.18 | 4.55 | 0.000 |
| Digit span | 3.55     | 0.90 | 3.95 | 0.000 |

$F(2,71) = 28.32$ ,  $p < 0.001$ , Adj. *R squared* = 0.43.



children with high exposure to Polish. An additional regression analysis conducted on 50% of bilingual children with highest weighted exposure and on matched monolingual peers revealed a main effect of Age ( $p < 0.01$ ), but only a marginally significant effect of Group ( $p = 0.05$ ), and no interaction of Group and Age ( $p = 0.89$ ).

## Productive Grammar Test (LITMUS-SRep)

When it comes to the productive grammar test, the bilingual group scored on average 76.12% out of 100 ( $SD = 17.48$ , range: 13.02–98.23), while the monolingual scores were close to ceiling ( $M = 90.80$  point,  $SD = 9.05$ , range: 60.18–99.79). The effect size as measured by Cohen's  $d$  was large ( $t(158) = 6.67$ ,  $p = 0.000$ , 95% CI [10.33, 19.03], *Cohen's  $d = 1.05$* ).

**Table 5** shows that the task results were predicted by the digit span, Raven scores, and L2 (English) cumulative exposure: the children with higher scores on STM and those with a higher IQ performed better on the SRep. However, the higher cumulative exposure to English resulted in the lower performance on the SRep test, as illustrated in **Figure 6**. There is a large gap in the performance on the task between the monolingual children and bilingual children with high exposure to English.

**TABLE 5 |** The best regression model predicting the productive grammar in the bilingual group.

|                                  | Estimate | SE   | <i>t</i> | <i>P</i> |
|----------------------------------|----------|------|----------|----------|
| Intercept                        | 38.71    | 6.94 | 5.58     | 0.000    |
| Raven                            | 0.58     | 0.26 | 2.20     | 0.031    |
| Digit span                       | 9.11     | 1.46 | 6.24     | 0.000    |
| L2 (English) cumulative exposure | −0.08    | 0.02 | −4.36    | 0.000    |

$F(3,76) = 25.75$ ,  $p < 0.001$ , *Adj. R squared* = 0.48.

The gap between the monolingual and bilingual children with low exposure to English is smaller. A regression analysis on 50% of bilingual children with the lowest weighted exposure to English and on matched monolingual peers revealed a significant effect of Age ( $p < 0.01$ ) and of Group ( $p < 0.001$ ), but the interaction between the two was non-significant ( $p > 0.7$ ).

## Phonological Processing Task (NWR)

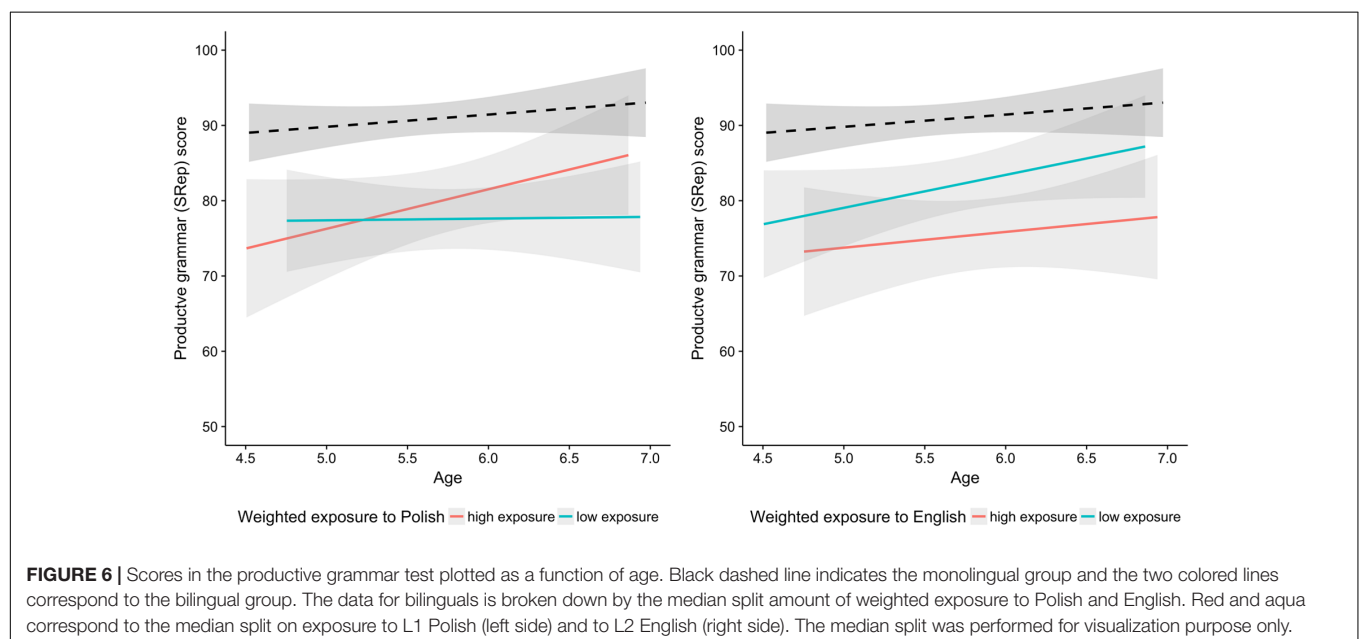
On the NWR task, the bilingual group scored on average 22.51 points out of 50 ( $SD = 9.23$ , range: 3–40) and the monolinguals scored 32.41 ( $SD = 8.06$ , range: 13–45). The effect size, as measured by Cohen's  $d$ , was large ( $t(156) = 7.18$ ,  $p = 0.000$ , 95% CI [7.17, 12.62], *Cohen's  $d = 1.14$* ).

**Table 6** shows that children with the higher digit span score and those with higher cumulative exposure to Polish had higher NWR scores. As indicated by **Figure 7**, the gap between the bilingual and monolingual children is lower for the bilinguals who had higher exposure to Polish. However, even for those children the gap does not seem to disappear with age, as also indicated by a regression analysis with 50% of bilinguals with highest weighted exposure to Polish and their matched monolingual peers. While there was a significant effect of Age ( $p < 0.01$ ) and Group ( $p < 0.001$ ), there was no significant interaction between them ( $p > 0.69$ ).

**TABLE 6 |** The best regression model predicting the phonological processing in the bilingual group.

|                                 | Estimate | SE   | <i>t</i> | <i>P</i> |
|---------------------------------|----------|------|----------|----------|
| Intercept                       | −2.88    | 4.76 | −0.60    | 0.547    |
| Digit span                      | 4.53     | 0.92 | 4.90     | 0.000    |
| L1 (Polish) cumulative exposure | 0.02     | 0.01 | 2.29     | 0.024    |

$F(2,76) = 15.56$ ,  $p < 0.001$ , *Adj. R squared* = 0.27.



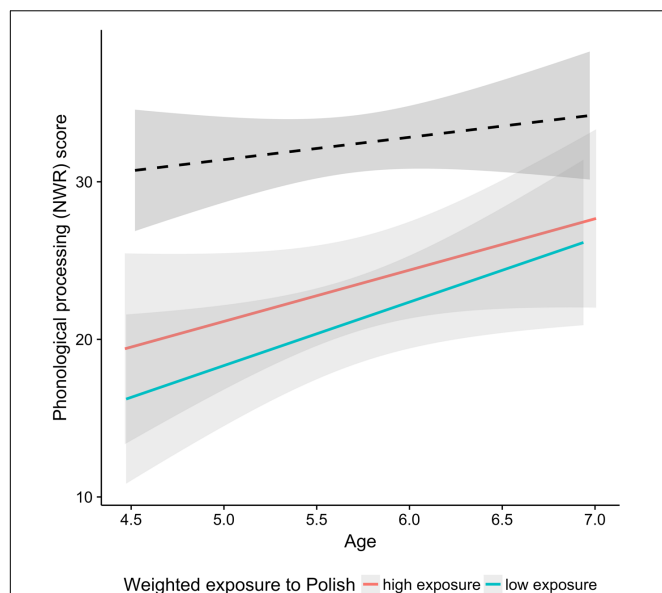


**FIGURE 7 |** Scores in the phonological processing test plotted as a function of age. The black dashed line indicates the monolingual group and the two colored lines correspond to the bilingual group. Red and aqua correspond to the median split on exposure to L1 Polish. The median split was performed for visualization purpose only.

## Discourse Task (LITMUS-MAIN)

In terms of the MAIN task, the bilingual group scored on average 8.13 points out of 17 for the story structure in the Telling condition ( $SD = 2.86$ , range: 1–16) and 9.21 points for the story

**TABLE 7 |** The best regression model predicting performance in the discourse task (story structure) in the bilingual group.

|                                  | Estimate | SE   | <i>t</i> | <i>P</i> |
|----------------------------------|----------|------|----------|----------|
| Intercept                        | 0.45     | 2.17 | 0.21     | 0.836    |
| L1 (Polish) cumulative exposure  | 0.02     | 0.00 | 4.01     | 0.000    |
| L2 (English) cumulative exposure | 0.01     | 0.01 | 2.38     | 0.021    |

$F(2,50) = 8.46$ ,  $p < 0.001$ , Adj.  $R$  squared = 0.22.

structure in the Retelling condition ( $SD = 3.18$ , range: 3–17). The monolingual group scored on average 7.36 points for the story structure in the Telling condition ( $SD = 2.71$ , range: 3–13) and 8.68 points for the story structure in the Retelling condition ( $SD = 2.98$ , range: 0–14). The Telling and Retelling scores correlated moderately in both groups (bilinguals:  $r = 0.45$ ,  $p = 0.001$ , monolinguals:  $r = 0.34$ ,  $p = 0.01$ ), therefore, for the further analyses we averaged the scores from the Telling and Retelling part of the task. When the scores were averaged, the bilingual group scored on average 8.63 points ( $SD = 2.57$ , range 2–15), while the monolinguals scored 8.02 points ( $SD = 2.33$ , range 1.5–12). The difference was not statistically significant, and the effect size as measured by Cohen's  $d$  was negligible ( $t(104) = 1.37$ ,  $p = 0.175$ , 95% CI  $[-0.29, 1.60]$ , Cohen's  $d = 0.27$ ).

Table 7 shows that the children with higher cumulative exposure to Polish and the children with higher cumulative exposure to English constructed more well-formed stories. This result is illustrated in Figure 8. The bilingual children with low exposure to Polish perform similarly to monolingual children on the task. The bilingual children with high exposure to Polish seem to score even higher than monolinguals.

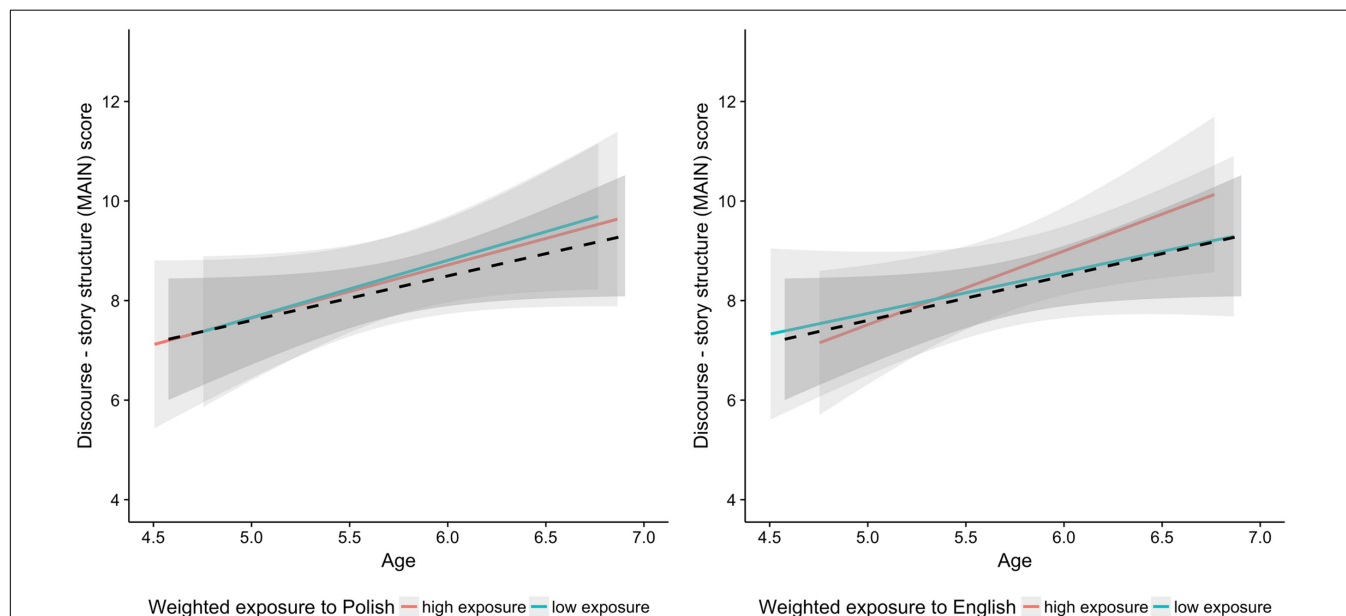


**FIGURE 8 |** Scores on the story structure plotted as a function of age. Black dashed line indicates the monolingual group and the two colored lines correspond to the bilingual group. The data for bilinguals is broken down by the median split amount of weighted exposure to Polish and English. Red and aqua correspond to the median split on exposure to L1 Polish (left side) and to L2 English (right side). The median split was performed for visualization purpose only.

## DISCUSSION

In this paper, we examined language skills in L1 Polish of Polish-English bilingual children (aged 4–7 years) growing up in the United Kingdom. We focused on four language domains: vocabulary (receptive and productive), grammar (receptive and productive), phonological processing, and discourse production (narration). We compared the overall scores in each task between bilinguals and monolinguals matched one-to-one on age, gender, maternal education, non-verbal IQ, and STM span. Further, in a series of regression analyses, we investigated the effect of cumulative exposure in L1 and L2 on the task scores, controlling for general cognitive abilities (non-verbal IQ and STM span), as well as SES and age. Finally, with another set of regression analyses we explored whether a greater amount of L1 exposure could possibly diminish the gap between the bilingual and monolingual children. Below, we first consider the results with regard to the overall performance of the bilinguals and monolinguals, and then focus on the contribution of language exposure to the language outcomes in the bilingual group.

### Differences between Bilingual Migrant Children and Their Monolingual Peers in the Four Domains of L1 Polish

The overall finding of our study is that in their performance on most L1 measures the bilinguals lagged behind their monolingual peers. There were large differences between the groups in terms of productive vocabulary, productive grammar (as measured by the SRep task), as well as phonological processing (as measured by the NWR task). There were also moderate differences between the groups in terms of receptive vocabulary and receptive grammar. However, the bilingual group did not differ from monolinguals in terms of story structure coherence in the narrative task. The results obtained are, to a large extent, consistent with the previous findings on L2 development in bilingual children.

With respect to the vocabulary size, previous research indicates that when tested in one language, bilingual children have smaller productive and receptive vocabulary than monolinguals (Pearson et al., 1993; O'Toole et al., 2017), even when tested in their L1 (e.g., Pearson et al., 1997; Uccelli and Páez, 2007; Miękisz et al., 2017). Our study adds new evidence to this body of research. It also provides new insights into identifying the sources of performance gap between bilinguals and monolinguals thanks to including a carefully matched monolingual control group. We have demonstrated that bilingual children have a smaller vocabulary in L1 than their monolingual peers even when their SES and general cognitive abilities are comparable.

The bilinguals also scored lower than monolinguals on both receptive and productive grammar tasks. This result replicates previous findings, showing developmental difficulties in L1 grammar among bilingual children (Thordardottir and Brandeker, 2013). We also observed that for the bilinguals (but not for the monolinguals) the productive grammar task was more difficult than the receptive grammar task. This result reflects the pattern that has been reported previously in studies on

L2 grammar performance in bilinguals. It shows that children struggle with the production of grammar, even if they have the receptive knowledge of the grammatical constructions tested (Chondrogianni and Marinis, 2012).

The large gap between the bilingual and monolingual children on the NWR task was more surprising, since many previous studies reported children scoring better on this phonological processing measure in L1 than in L2 (Gutiérrez-Clellen and Simon-Cerejido, 2010; Summers et al., 2010). However, the NWR task used in our study might be more sensitive to problems with L1 phonological processing, since it deliberately contained many phonological structures typical for Polish. This might have resulted in the effect obtained for bilinguals on NWR, in contrast to previous research, which utilized various types of quasi-universal tasks. Delays in L1 phonological development among bilingual children have been reported before, which makes this explanation plausible (Goldstein and Washington, 2001; Fabiano-Smith and Barlow, 2010).

As far as the discursive abilities are concerned, in the MAIN task, the bilinguals scored on par with their monolingual peers for the story structure of their narratives, which replicated the finding by Kunnari et al. (2016). This result can be explained by the fact that the narrative abilities intersect children's language abilities and their pragmatic awareness (Reese et al., 2010). Telling a coherent narrative requires robust cognitive skills necessary for building a logical storyline, so children's discourse abilities probably go beyond their language-specific skills (Paradis et al., 2014; Gagarina et al., 2016). Previous studies have shown that similar age-dependent narrative patterns are shared by monolingual children from different language backgrounds (Berman and Slobin, 1994) and it seems that narrative abilities develop similarly in bilingual and monolingual children.

### The Impact of the Cumulative Exposure to L1 and L2 on Language Performance across the Four Domains

The second set of findings relates to the effects of exposure on language measures. We have found that the cumulative exposure to L1 was related to higher scores in the receptive vocabulary, productive vocabulary, phonological processing, and discourse. We also found an adverse effect of L2 cumulative exposure on only one language measure – the productive grammar, and its seemingly surprising positive effect on the narrative production. For the receptive grammar, we found no significant effect of exposure to L1 or to L2 once other factors have been controlled for.

Overall, the results suggest that language exposure is crucial primarily for the productive tasks (producing grammar and vocabulary, repeating non-words and producing narratives) and has less of an impact on the comprehension tasks. This finding is in line with the previous research on bilingual children that shows the influence of language exposure on productive tasks in L1 (Thordardottir and Brandeker, 2013 – SRep; Patterson, 2002 – vocabulary; Summers et al., 2010 – NWR). Moreover, it aligns with an earlier study by Thordardottir (2011), who found that although language exposure influenced both the

receptive and productive vocabulary scores of bilingual children, the effect was much greater for the productive tasks. While the complete lack of effect of exposure on the receptive grammar tasks contradicts previous research (Gathercole and Thomas, 2009; Thordardottir, 2011), this discrepancy might be due to the fact that previous studies did not fully control for the factors related to the cognitive development of children, such as the non-verbal IQ and STM span. In our study, the two factors strongly predicted the receptive grammar scores, and these general cognitive abilities explained most of the variance in this language task.

The differential effects of the cumulative exposure on the receptive and productive tasks found in the current analysis are of vital importance, because they suggest that exposure to the home language is critical for mastering the productive skills in this language. It also appears that the performance in the receptive tasks is much less impacted by the amount of language exposure: it is easier to understand than to produce language having had little exposure to that language. It is also worth adding that the inter-subject variability in the receptive grammar performance was much smaller than in the production task and so was the performance gap between monolinguals and bilinguals.

Another issue is the negative effect of L2 exposure on the L1 production of grammatical structures in the SRep task. This result suggests the existence of negative transfer from L2 to the L1 (Paradis and Navarro, 2003; Bernardini and Schlyter, 2004; Kupisch, 2007). More specifically, when repeating Polish sentences in the SRep task, which involves accessing the mental representation of a given structure, the knowledge of English syntactic templates possibly interfered with the knowledge of Polish syntax, leading to errors in the production of syntactically complex Polish sentences. Another possible explanation is that the early acquisition of English, a language less morphosyntactically complex than Polish, “desensitized” children to the complexity of Polish inflection (van der Slik et al., 2017). However, at this point the above interpretations of the negative effect of L2 exposure on the scores in L1 SRep task are only speculative. A qualitative error analysis would be required to determine the precise sources of difficulty in SRep. Thus, further research is needed to determine in what ways L2 exposure may affect L1 grammatical performance.

A separate question is why there was no impact of exposure to L1 Polish on the performance in the SRep task. One hypothetical explanation is purely statistical: if the indices of exposure to L1 and to L2 were highly collinear, introducing one of the indices might have “pushed out” the other from the model. However, in this case the two indices share little common variance (14%), so this explanation is rather unlikely. It is thus more plausible that the L1 input typically directed to bilingual children in the migrant context does not systematically familiarize them with the syntactic knowledge required to repeat more complex sentences (e.g., object and subject relative clauses, conditionals, object and subject clefts and noun complement clauses). Hence, the large variability in the SRep scores in the bilingual children and the absence of any impact of L1 exposure. In contrast, the monolingual children might systematically be exposed to such

structures not only at home, but also in educational settings and through the media, which would explain why their scores were higher and less varied. However, more research is needed on the features of home discourse and its relationship with children’s syntactic development to further substantiate this claim.

At the same time, there is a positive influence of L2 exposure on the discourse production, as the bilingual children’s narrative abilities are positively correlated with both L1 and L2 exposure. This is consistent with the previous research, which suggests that the ability to create coherent stories is independent of the language-specific skills (e.g., Iluz-Cohen and Walters, 2012; Gagarina, 2016; Rodina, 2016). If the ability to tell coherent stories is less reliant on the specific language skills, but depends on the child’s pragmatic awareness (Reese et al., 2010), such an awareness develops in contact with any of the two languages of the bilingual child. Possibly, in the initial years of schooling, there is a carry-over of the child’s narrative abilities across the two languages, even if the child’s linguistic abilities in one of the languages are weaker (Gagarina, 2016). This suggests that in child bilinguals, exposure to any language builds a language-universal ability to structure stories in a coherent way.

## Can High Exposure to L1 Minimize the Potential Gap between Monolinguals and Bilinguals?

While the high exposure to L1 positively influenced language outcomes in the bilingual children, our study suggests that it might not be enough to minimize the gap between the monolingual and bilingual children. For each task in which we observed the effect of cumulative exposure to L1, we conducted an additional analysis on the bilingual children with the highest weighted L1 exposure and a matched monolingual group. In the analyses, we tested for an interaction between group and age on the task outcomes. The presence of such an interaction would indicate that with age, the bilingual children with high rates of L1 exposure “catch up” with their monolingual peers. Contrary to our expectations, we have not observed such an effect for any of the tasks where a performance gap was observed. It is interesting to note that this result was consistent across the board – i.e., for grammatical, lexical and phonological tasks, both productive and receptive. This indicates that even though exposure might be more crucial for productive rather than receptive tasks, there is no domain in which high exposure to L1 guarantees outcomes comparable to that of monolingual peers. Overall, the results suggest that although exposing bilingual children to L1 will certainly benefit their L1 language performance, it might not be enough to minimize gaps in L1 skills between them and their monolingual peers.

## CONCLUSION

Our study shows that when tested in L1, bilingual children lag behind their monolingual peers on vocabulary, grammar, and phonological processing. The performance differences between the two groups are most prominent in the productive language

tasks. At the same time, the productive tasks are also more influenced by cumulative language exposure than the receptive tasks. While high exposure to L1 might not be enough to close the performance gap between the bilingual and monolingual children, providing exposure to L1 and promoting situations in which bilinguals could practice their production in L1 will certainly benefit their development in that language. This finding is essential not only for furthering our understanding of bilingual language acquisition, but it also has important practical implications. Unlike many other migrant communities who stay in the host country for good, Polish families often re-migrate to Poland, the home country of the parents. Upon returning to their home country, many children of Polish migrants experience educational setbacks due to inferior knowledge of their L1 Polish, as compared to their monolingual peers (Grzymała-Moszczyńska et al., 2015). Our study points to the areas where these children might experience most difficulties – namely productive vocabulary and grammar. It also shows that extensive and varied exposure to L1 in these areas would certainly be beneficial. These clues might be used to design better interventions for the migrant children who return to their home countries and who face language difficulties in their L1. It also shows that narration (story structure) is a strength in bilinguals' L1 performance. Hence the interventions could build on this strength when trying to enhance vocabulary and grammar.

The results of the current study provide support for the claims made in the heritage language literature that the L1 of young heritage language speakers resembles more an L2 learned in the adulthood, than an L1 naturally acquired in the childhood (e.g., Rothman, 2009). Heritage speakers usually end up being dominant in the majority language in adulthood, no matter whether both their languages are present from birth, or the majority language is acquired later due to migration in childhood (Cabo and Rothman, 2012). The results reported in the current study suggest that the Polish-English bilinguals growing up in the United Kingdom may never reach the level of a monolingual Polish peers growing up in Poland. Such “incomplete L1 acquisition” is defined by Montrul (2008, p. 21), as “a mature linguistic state, the outcome of language acquisition that is not complete (...) in childhood (...), when some specific properties of the language do not have a chance to reach age-appropriate levels of proficiency after intense exposure to the L2 begins.” We speculate that such a scenario is likely, if the Polish-English bilinguals stay in the United Kingdom for good and maintain only sporadic contacts with their Polish-speaking families in Poland.

Our study is not without limitations. First, our bilingual sample is not fully representative of the population of Polish migrant children in the United Kingdom. The families who took part in our research were volunteers, which means they were possibly interested in the subject of bilingualism (Haman et al., 2014). Thus, our data set did not include a large part of the Polish migrant population in the United Kingdom who do not support maintaining L1 in their children (e.g., for the sake of acculturation and integration with the ambient society). It is therefore likely that our data paint an overly optimistic picture

of the L1 performance in bilingual child migrants. The second limitation is that the current report includes only the analyses of L1 performance, but no analogous analyses L2 performance, which will be completed in the near future. The last limitation is that both languages of the bilingual children are Indo-European and both follow the canonical SVO word order. It is thus not clear how our findings can translate to pairs of more typologically distinct languages.

Nevertheless, the reported study is unique in that it presents a comprehensive analysis of bilingual children's L1 across a range of language domains. An additional value of the study is that the performance of the bilingual migrant children was compared with that of carefully matched monolinguals. We were also able to isolate the impact of language exposure in both L1 and L2 on language skills, while at the same time controlling for a range of other factors known to contribute to performance in language tasks. In the future, we plan to extend our exploration by conducting more detailed error analyses on the collected language material. This should reveal the most problematic areas which account for the gap between monolinguals and bilinguals in the domains of vocabulary, grammar, and phonological processing. In particular, the analysis of errors in sentence comprehension and production should be valuable as it will shed more light on the issue of cross-linguistic influence between the two languages of a bilingual.

## ETHICS STATEMENT

Study presented in this paper was approved by the Komisja ds. Etyki Badań Naukowych (Ethics Committee) at the Faculty of Psychology, University of Warsaw. Parents of all children who participated in the study signed informed consent forms. Children gave oral assent to take part in the study.

## AUTHOR CONTRIBUTIONS

Leading the main project: EH and ZW (both first authors). Leading the supporting projects: ZW and AO. Conception and design of the study: EH, ZW, JS, MB-P, AO, and AM. Design of tasks used in the study: EH, ZW, JS, MB-P, AO, and NB. Data collection: KM, JK, AK, AM, MF-N. Data coding and analysis: EH, ZW, MM, JS, MB-P, AO, KM, MŁ, JK, AM, AK, NB, and MF-N. Data interpretation: EH, ZW, MM, JS, MB-P, AO, KM, MŁ, JK, AM, AK, NB, and MF-N. Drafting the article: MM and JS. Critical revision and rewriting of the article: EH, ZW, MM, JS, MB-P, AO, KM, MŁ, JK, AM, AK, NB, and MF-N. Final approval of the version to be published: EH, ZW, MM, JS, MB-P, AO, KM, MŁ, JK, AM, AK, NB, and MF-N.

## FUNDING

The data for this paper come from the Bi-SLI-Poland project entitled “Cognitive and language development of Polish bilingual children at the school entrance age – risks and opportunities”

conducted within the European COST Action IS0804 “Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment” and carried out at the Faculty of Psychology, University of Warsaw, Poland in collaboration with Institute of Psychology, Jagiellonian University, Poland. The project was supported by the Polish Ministry of Science and Higher Education/National Science Centre (Decision 809/N-COST/2010/0). Data collection and coding were also partly supported by the Polish Ministry of Science and Higher Education grant (0094/NPRH3/H12/82/2014) Phonological and Morpho-syntactic Features of Language and Discourse of Polish Children Raised Bilingually in Migrant Communities in Great Britain carried out at the Faculty of Modern Languages, University of Warsaw, Poland, and Foundation for Polish Science subsidy to ZW.

## REFERENCES

- Aoyama, K., Guion, S. G., Flege, J. E., and Yamada, T. (2008). The first years in an L2-speaking environment: a comparison of Japanese children and adults learning American English. *Int. Rev. Appl. Linguist. Lang. Teach.* 46, 61–90. doi: 10.1515/IRAL.2008.003
- Armon-Lotem, S. (2016). Disentangling bilingualism from SLI: dissociating exposure and input. *Bilingualism* 20, 33–34. doi: 10.1017/S1366728916000377
- Armon-Lotem, S., de Jong, J., and Meir, N. (eds). (2015). *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual Matters.
- Asher, J. J., and García, R. (1969). The optimal age to learn a foreign language. *Mod. Lang. J.* 53, 334–341. doi: 10.1111/j.1540-4781.1969.tb04603.x
- Au, T. K.-F., Knightly, L. M., Jun, S.-A., and Oh, J. S. (2002). Overhearing a language during childhood. *Psychol. Sci.* 13, 238–243. doi: 10.1111/1467-9280.00444
- Barlow, J. A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Front. Psychol.* 5:288. doi: 10.3389/fpsyg.2014.00288
- Bedore, L. M., Pena, E. D., Griffin, Z. M., and Hixon, J. G. (2016). Effects of age of English Exposure, Current Input/Output, and grade on bilingual language performance. *J. Child Lang.* 43, 687–706. doi: 10.1017/S0305000915000811
- Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: opportunities and challenges for linguistics. *Theor. Linguist.* 39, 129–181. doi: 10.1515/tl-2013-0009
- Berman, R. A., and Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. London, England: Routledge.
- Bernardini, P., and Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: the Ivy Hypothesis. *Bilingualism* 7, 49–69. doi: 10.1017/S1366728904001270
- Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism* 13, 525–531. doi: 10.1017/S1366728909990423
- Bishop, D. V. M. (2003). *Test for Reception of Grammar: Version 2: TROG-2*. London: Pearson Assessment.
- Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., and Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *J. Speech Lang. Hear. Res.* 58, 1747–1760. doi: 10.1044/2015\_JSLHR-L-15-0058
- Bohnacker, U. (2016). Tell me a story in English or Swedish: narrative production and comprehension in bilingual preschoolers and first graders. *Appl. Psycholinguist.* 37, 19–48. doi: 10.1017/S0142716415000405
- Bolonyai, A. (2007). (In)vulnerable agreement in incomplete bilingual L1 learners. *Int. J. Biling.* 11, 3–23. doi: 10.1177/13670069070110010201
- Botting, N. (2005). Non-verbal cognitive development and language impairment. *J. Child Psychiatry* 46, 317–326. doi: 10.1111/j.1469-7610.2004.00355.x
- Cabo, D. P. Y., and Rothman, J. (2012). The (il) logical problem of heritage speaker bilingualism and incomplete acquisition. *Appl. Linguist.* 3, 450–455. doi: 10.1093/applin/ams037
- Carroll, S. E. (2017). Explaining bilingual learning outcomes in terms of exposure and input. *Bilingualism* 20, 37–41. doi: 10.1017/S1366728916000511
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, morphology and complex syntax in successive bilingual children. *Linguist. Approaches Biling.* 1, 318–345. doi: 10.1075/lab.1.3.05cho
- Chondrogianni, V., and Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism* 15, 5–21. doi: 10.1017/S1366728911000368
- Conboy, B. T., and Thal, D. J. (2006). Ties between the lexicon and grammar: cross-sectional and longitudinal studies of bilingual toddlers. *Child Dev.* 77, 712–735. doi: 10.1111/j.1467-8624.2006.00899.x
- De Houwer, A. (2005). “Early bilingual acquisition: focus on morphosyntax and the separate development hypothesis,” in *The Handbook of Bilingualism*, eds J. F. Kroll and A. M. B. De Groot (Oxford: Oxford University Press), 30–48.
- De Houwer, A. (2007). Parental language input patterns and children’s bilingual use. *Appl. Psycholinguist.* 28, 411–424. doi: 10.1017/S0142716407070221
- De Houwer, A. (2015). Harmonious bilingual development: young families’ well-being in language contact situations. *Int. J. Biling.* 19, 169–184. doi: 10.1177/1367006913489202
- Dodd, B., Zhu, H., Crosbie, S., Holm, A., and Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London: Psychology Corporation.
- Döpke, S. (1998). Competing language structures: the acquisition of verb placement by bilingual German-English children. *J. Child Lang.* 25, 555–584. doi: 10.1017/S0305000998003584
- Ebert, K. D., and Kohnert, K. (2016). Language learning impairment in sequential bilingual children. *Lang. Teach.* 49, 301–338. doi: 10.1017/S0261444816000070
- En, L. G. W., Brebner, C., and McCormack, P. (2014). A preliminary report on the English phonology of typically developing English–Mandarin bilingual preschool Singaporean children. *Int. J. Lang. Commun. Disord.* 49, 317–332. doi: 10.1111/1460-6984.12075
- Fabiano-Smith, L., and Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: evidence from phonetic inventories. *Int. J. Biling. Educ. Biling.* 13, 81–97. doi: 10.1080/13670050902783528
- Fabiano-Smith, L., and Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish–English speaking children. *J. Speech Lang. Hear. Res.* 53, 160–178. doi: 10.1044/1092-4388(2009/07-0064)
- Fiestas, C. E., and Peña, E. D. (2004). Narrative discourse in bilingual children: language and task effects. *Lang. Speech Hear. Serv. Sch.* 35, 155–168. doi: 10.1044/0161-1461(2004/016)

## ACKNOWLEDGMENTS

We express our gratitude to all children and parents who participated in the study, as well as to teachers in preschools who helped in conducting the study and to all research assistants and volunteers who helped with testing and data coding. We thank previous members of our team who contributed their expertise at various stages of the project: Magdalena Karwala, Dorota Kiebzak-Mandera, Katarzyna Kuś, and Anna Marzecova. We would also like to thank Magdalena Smoczyńska for her inspiration during the initial stages of the project and Elin Thordardottir for her insightful comments on the first version of this paper. Last but not least, we would like to thank Theo Marinis and Shula Chiat for their valuable input in the development of the tasks and support in organizing data collection in the United Kingdom.

- Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience*, ed. W. Strange (Timonium, MD: York Press), 233–277.
- Flege, J. E. (2002). "Interactions between the native and second-language phonetic systems," in *An Integrated View of Language Development. Papers in Honor of Henning Wode*, eds P. Burmeister, T. Piske, and A. Rhode (Trier: Wissenschaftlicher Verlag), 217–244.
- Flege, J. E., and Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *J. Acoust. Soc. Am.* 91, 370–389. doi: 10.1121/1.402780
- Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: rasskaz and Erzählung. *Appl. Psycholinguist.* 37, 91–122. doi: 10.1017/S0142716415000430
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., et al. (2012). *MAIN: Multilingual Assessment Instrument for Narratives*, Vol. 56. Berlin: ZAS.
- Gagarina, N., Klop, D., Tsimpli, I. M., and Walters, J. (2016). Narrative abilities in bilingual children. *Appl. Psycholinguist.* 37, 11–17. doi: 10.1017/S0142716415000399
- Gathercole, S. E., Willis, C. S., Emslie, H., and Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: a longitudinal study. *Dev. Psychol.* 28, 887–898. doi: 10.1037/0012-1649.28.5.887
- Gathercole, V. M. (2002a). "Grammatical gender in bilingual and monolingual children: a Spanish morphosyntactic distinction," in *Language and Literacy in Bilingual Children*, eds K. D. Oller and R. Eilers (Clevedon: Multilingual Matters), 207–219.
- Gathercole, V. M. (2002b). "Monolingual and bilingual acquisition: learning different treatments of that-trace phenomena in English and Spanish," in *Language and Literacy in Bilingual Children*, eds K. D. Oller and R. Eilers (Clevedon: Multilingual Matters), 220–254.
- Gathercole, V. M. (ed.). (2013a). *Issues in the Assessment of Bilinguals*. Bristol: Multilingual Matters.
- Gathercole, V. M. (ed.). (2013b). *Solutions for the Assessment of Bilinguals*. Bristol: Multilingual Matters.
- Gathercole, V. M., and Thomas, E. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingualism* 12, 213–237. doi: 10.1017/S1366728909004015
- Genesee, F., and Nicoladis, E. (2007). "Bilingual first language acquisition," in *Blackwell Handbook of Language Development*, eds E. Hoff and M. Shatz (Oxford: Blackwell), 324–342. doi: 10.1002/9780470757833.ch16
- Godson, L. (2004). Vowel production in the speech of Western Armenian heritage speakers. *Herit. Lang. J.* 2, 45–70.
- Goldstein, B., and Washington, P. S. (2001). An initial investigation of phonological patterns in typically developing 4-year-old Spanish-English bilingual children. *Lang. Speech Hear. Serv. Sch.* 32, 153–164. doi: 10.1044/0161-1461(2001/014)
- Grzymała-Moszczyńska, H., Grzymała-Moszczyńska, J., Durlik, J., and Szydłowska, P. (2015). *(Nie)łatwe powroty do domu? Funkcjonowanie dzieci i młodzieży powracających z emigracji*. Warsaw: Fundacja Centrum im. prof. Bronisława Gereńki.
- Gutiérrez-Clellen, V. F., and Simon-Cereijido, G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English speaking children: Does the language of assessment matter? *Learn. Disabil. Res. Pract.* 25, 48–58. doi: 10.1111/j.1540-5826.2009.00300.x
- Gutiérrez-Clellen, V. F., Simon-Cereijido, G., and Wagner, C. (2008). Bilingual children with language impairment: a comparison with monolinguals and second language learners. *Appl. Psycholinguist.* 29, 3–19. doi: 10.1017/S0142716408080016
- Güven, C., and Islam, A. (2015). Age at migration, language proficiency, and socioeconomic outcomes: evidence from Australia. *Demography* 52, 513–542. doi: 10.1007/s13524-015-0373-6
- Haman, E., and Fronczyk, K. (2012). *Obrazkowy Test Słownikowy - Rozumienie (OTSR)*. Gdańsk: Pracownia Testów Psychologicznych i Pedagogicznych.
- Haman, E., Fronczyk, K., and Łuniewska, M. (2012). *Obrazkowy Test Słownikowy - Rozumienie (OTSR)*. Podręcznik: Pracownia Testów Psychologicznych i Pedagogicznych.
- Haman, E., Wodniecka, Z., Kołak, J., Łuniewska, M., and Mieszkowska, K. (2014). "Social aspects of psycholinguistic research: reflections on the ongoing study of cognitive and linguistic development of Polish immigrant Children," in *Zweisprachigkeit und Bilingualer Unterricht*, eds M. Olpińska-Szkielko and L. Bartelle (Frankfurt: Peter Lang), 77–86.
- Hoff, E. (2006). How social contexts support and shape language development. *Dev. Rev.* 26, 55–88. doi: 10.1016/j.dr.2005.11.002
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev. Psychol.* 49, 4–14. doi: 10.1037/a0027238
- Hoff, E., and Core, C. (2013). Input and language development in bilingually developing children. *Semin. Speech Lang.* 34, 215–226. doi: 10.1055/s-0033-1353448
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Hoff, E., Quinn, J. M., and Giguere, D. (2017). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Dev. Sci.* doi: 10.1111/desc.12536 [Epub ahead of print].
- Iluz-Cohen, P., and Walters, J. (2012). Telling stories in two languages: narratives of bilingual preschool children with typical and impaired language. *Bilingualism* 15, 58–74. doi: 10.1017/S1366728911000538
- Isurin, L., and Ivanova-Sullivan, T. (2008). Lost in between: the case of Russian heritage speakers. *Herit. Lang. J.* 6, 72–104.
- Jaworowska, A., and Szustrowa, T. (2003). *Test Matryc Ravena – Wersja Kolorowa*. Warsaw: Pracownia Testów Psychologicznych PTP.
- Jones, G. (2011). A computational simulation of children's performance across three nonword repetition tests. *Cogn. Syst. Res.* 12, 113–121. doi: 10.1016/j.cogsys.2010.07.008
- Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., and Pine, J. M. (2010). Lexicality and frequency in specific language impairment: accuracy and error data from two nonword repetition tests. *J. Speech Lang. Hear. Res.* 53, 1642–1655. doi: 10.1044/1092-4388(2010/09-0222)
- Kail, R. (2000). Speed of information processing: developmental change and links to intelligence. *J. Sch. Psychol.* 38, 51–61. doi: 10.1016/S0022-4405(99)00036-9
- Kapalková, S., Polišínská, K., Markova, L., and Fenton, J. (2016). Narrative abilities in early successive bilingual Slovak–English children: a cross-language comparison. *Appl. Psycholinguist.* 37, 145–164. doi: 10.1017/S0142716415000454
- Kehoe, M. (2002). Developing vowel systems as a window to bilingual phonology. *Int. J. Biling.* 6, 315–334. doi: 10.1177/1367069020060030601
- Kiebzak-Mandera, D., Otwinowska, A., and Białecka-Pikul, M. (2012). *MAIN Multilingual Assessment Instrument for Narratives: Polish Version*. ZAS Papers in Linguistics, 56. Berlin: ZAS.
- Kim, J.-H., Montrul, S., and Yoon, J. (2009). Binding interpretation of anaphors in Korean heritage speakers. *Lang. Acquis.* 16, 3–35. doi: 10.1080/10489220802575293
- Kohnert, K., Windsor, J., and Ebert, K. D. (2009). Primary or "specific" language impairment and children learning a second language. *Brain Lang.* 109, 101–111. doi: 10.1016/j.bandl.2008.01.009
- Kohnert, K., Windsor, J., and Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learn. Disabil. Res. Pract.* 21, 19–29. doi: 10.1111/j.1540-5826.2006.00204.x
- Komeili, M., and Marshall, C. R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clin. Linguist. Phon.* 27, 152–162. doi: 10.3109/02699206.2012.751625
- Kułańska, M. (2013). *Polski uczeń odbiorcą brytyjskich polityk edukacyjnych – kontekst instytucjonalny*. Paper Presented at the EuroEmgranci. PL Conference, Cracow.
- Kunnari, S., Välimaa, T., and Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual Finnish and bilingual Finnish–Swedish children. *Appl. Psycholinguist.* 37, 123–144. doi: 10.1017/S0142716415000442
- Kupisch, T. (2007). Determiners in bilingual German–Italian children: what they tell us about the relation between language influence and language dominance. *Bilingualism* 10, 57–78. doi: 10.1017/S1366728906002823
- Leseman, P. P. M. (2010). Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *J. Multiling. Multicult. Dev.* 21, 93–112. doi: 10.1080/01434630008666396

- Lleó, C. (2002). The role of markedness in the acquisition of complex prosodic structures by German-Spanish bilinguals. *Int. J. Biling.* 6, 291–313. doi: 10.1177/13670069020060030501
- Lumley, T., and Miller, A. (2004). *Leaps: Regression Subset Selection. R Package Version, 2*.
- Marecka, M., Wrembel, M., Zembrzski, D., and Otwinowska-Kasztelanica, A. (2016). “Do early bilinguals speak differently than their monolingual peers? Predictors of phonological performance of Polish-English bilingual children,” in *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015*, eds E. Babatsouli and D. Ingram (Chania: WA Publications), 207–213.
- Marinis, T., and Armon-Lotem, S. (2015). “Sentence repetition,” in *Assessing Multilingual Children: disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. De Jong, and N. Meir (Clevedon: Multilingual matters), 95–121.
- Marinis, T., Chiat, S., Armon-Lotem, S., Gibbons, D., and Gipps, E. (2010). *School-Age Sentence Imitation Test (SASIT)*. Reading: University of Reading.
- Maviş, İ., Tunçer, M., and Gagarina, N. (2016). Macrostructure components in narrations of Turkish-German bilingual children. *Appl. Psycholinguist.* 37, 69–89. doi: 10.1017/S0142716415000429
- Mayr, R., Howells, G., and Lewis, R. (2014). Asymmetries in phonological development: the case of word-final cluster acquisition in Welsh-English bilingual children. *J. Child Lang.* 42, 146–179. doi: 10.1017/S0305000913000603
- Miękisz, A., Haman, E., Łuniewska, M., Kuś, K., O’Toole, C., and Katsos, N. (2017). The impact of a first-generation immigrant environment on the heritage language: productive vocabularies of Polish toddlers living in the UK and Ireland. *Int. J. Biling. Educ. Biling.* 20, 183–200. doi: 10.1080/13670050.2016.1179259
- Mieszkowska, K., Łuniewska, M., Kołak, J., Kacprzak, A., Wodniecka, Z., and Haman, E. (2017). Home language will not take care of itself: vocabulary knowledge in trilingual children in the United Kingdom. *Front. Psychol.* 8:1358. doi: 10.3389/fpsyg.2017.01358
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: John Benjamins. doi: 10.1075/sibil.39
- Montrul, S. (2010). Current issues in heritage language acquisition. *Annu. Rev. Appl. Linguist.* 30, 2–23. doi: 10.1017/S0267190510000103
- Montrul, S., and Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism* 13, 449–473. doi: 10.1017/S1366728910000040
- Montrul, S., and Potowski, K. (2007). Command of gender agreement in school-age Spanish-English Bilingual Children. *Int. J. Biling.* 11, 301–328. doi: 10.1177/13670069070110030301
- Muñoz, M. L., Gillam, R. B., Peña, E. D., and Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latin children. *Lang. Speech Hear. Serv. Sch.* 34, 332–342. doi: 10.1044/0161-1461(2003/027)
- Nagy, N., and Kochetov, A. (2013). “Voice onset time across the generations: a cross-linguistic study of contact-induced change,” in *Multilingualism and Language Contact in Urban Areas: Acquisition—Development—Teaching—Communication*, eds P. Siemund, I. Gogolin, M. E. Schulz, and J. Davydova (Amsterdam: John Benjamins Publishing), 19–38.
- Office for National Statistics [ONS] (2014). *Migration Statistics Quarterly Report*. Available at: [http://www.ons.gov.uk/dcp171778\\_386531.pdf](http://www.ons.gov.uk/dcp171778_386531.pdf)
- Oh, J. S., Jun, S.-A., Knightly, L. M., and Au, T. K.-F. (2003). Holding on to childhood language memory. *Cognition* 86, B53–B64. doi: 10.1016/S0010-0277(02)00175-0
- Okólski, M., and Salt, J. (2014). Polish emigration to the UK after 2004; why did so many come? *Cent. East. Eur. Migr. Rev.* 3, 11–37.
- O’Toole, C., Gatt, D., Hickey, T. M., Haman, E., Miękisz, A., Armon-Lotem, S., et al. (2017). Parent report of early lexical production in bilingual children across varied contexts: a CDI study. *Int. J. Biling. Educ. Biling.* 20, 124–145. doi: 10.1080/13670050.2016.1179258
- Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *Int. J. Biling.* 5, 19–38. doi: 10.1177/13670069010050010201
- Paradis, J. (2009). “Second language acquisition in children,” in *Blackwell Handbook of Language Development*, eds E. Hoff and M. Shatz (Oxford: Blackwell Publishing), 62.
- Paradis, J., and Genesee, F. (1996). Syntactic acquisition in bilingual children. *Stud. Second Lang. Acquis.* 18, 1–25. doi: 10.1017/S0272263100014662
- Paradis, J., Genesee, F., and Crago, M. B. (2014). “Dual language development and disorders,” in *A Handbook on Bilingualism and Second Language Learning*, eds J. Paradis, F. Genesee, and M. B. Crago (Baltimore, MD: Brookes Publishing).
- Paradis, J., and Kirova, A. (2014). English second-language learners in preschool: profile effects in their English abilities and the role of home language environment. *Int. J. Behav. Dev.* 38, 342–349. doi: 10.1177/0165025414530630
- Paradis, J., and Navarro, S. (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *J. Child Lang.* 30, 371–393. doi: 10.1017/S0305000903005609
- Parra, M., Hoff, E., and Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *J. Exp. Child Psychol.* 108, 113–125. doi: 10.1016/j.jecp.2010.07.011
- Patterson, J. L. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Appl. Psycholinguist.* 23, 493–508. doi: 10.1017/S0142716402004010
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Appl. Psycholinguist.* 18, 41–58. doi: 10.1017/S0142716400009863
- Pearson, B. Z., Fernández, S. C., and Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43, 93–120. doi: 10.1111/j.1467-1770.1993.tb00174.x
- Polinsky, M. (2006). Incomplete acquisition: American Russian. *J. Slavic Linguist.* 14, 191–262.
- Polinsky, M. (2008). Russian gender under incomplete acquisition. *Herit. Lang. J.* 6, 40–71.
- Polinsky, M. (2011). Reanalysis in adult heritage language: new evidence in support of attrition. *Stud. Second Lang. Acquis.* 33, 305–328. doi: 10.1017/S027226311000077X
- Polinsky, M. (2016). Bilingual children and adult heritage speakers: the range of comparison. *Int. J. Biling.* doi: 10.1177/1367006916656048
- Portes, A., and Hao, L. (1998). E Pluribus Unum: bilingualism and language loss in the second generation. *Sociol. Educ.* 71, 269–294. doi: 10.2307/2673171
- Potowski, K. (2013). “Language maintenance and shift,” in *The Oxford Handbook of Sociolinguistics*, eds R. Bayley, R. Cameron, and C. Lucas (Oxford: Oxford University Press).
- Qi, C. H., Kaiser, A. P., Milan, S., and Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Lang. Speech Hear. Serv. Sch.* 37, 5–16. doi: 10.1044/0161-1461(2006/002)
- Raven, J. C. (2003). *Raven’s Coloured Progressive Matrices (CPM)*. Cambridge: Pearson Publishing.
- Reese, E., Leyva, D., Sparks, A., and Grolnick, W. (2010). Maternal elaborative reminiscing increases low-income children’s narrative skills relative to dialogic reading. *Early Educ. Dev.* 21, 318–342. doi: 10.1080/10409289.2010.481552
- Roch, M., Florit, E., and Levorato, C. (2016). Narrative competence of Italian-English bilingual children between 5 and 7 years. *Appl. Psycholinguist.* 37, 49–67. doi: 10.1017/S0142716415000417
- Rodina, Y. (2016). Narrative abilities of preschool bilingual Norwegian-Russian children. *Int. J. Biling.* doi: 10.1177/1367006916643528
- Rodriguez, J. L., Diaz, R. M., Duran, D., and Espinosa, L. (1995). The impact of bilingual preschool education on the language development of Spanish-speaking children. *Early Child. Res. Q.* 10, 475–490. doi: 10.1016/0885-2006(95)90017-9
- Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: inflected infinitives in Heritage Brazilian Portuguese. *Int. J. Biling.* 11, 359–389. doi: 10.1177/13670069070110040201
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: romance languages as heritage languages. *Int. J. Biling.* 13, 155–163. doi: 10.1177/1367006909339814
- Rothman, J., and Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Appl. Linguist.* 35, 93–98. doi: 10.1093/applin/amt049
- Schwartz, M. (2008). Exploring the relationship between family language policy and heritage language knowledge among second generation Russian-Jewish immigrants in Israel. *J. Multiling. Multicult. Dev.* 29, 400–418. doi: 10.1080/01434630802147916

- Scontras, G., Fuchs, Z., and Polinsky, M. (2015). Heritage language and linguistic theory. *Front. Psychol.* 6:1545. doi: 10.3389/fpsyg.2015.01545
- Shields, M. A., and Price, S. W. (2004). "The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas," in *How Labor Migrants Fare*, eds P. D. K. F. Zimmermann and D. A. Constant (Berlin: Springer), 167–190.
- Snow, C. E., and Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Lang. Speech* 20, 357–365. doi: 10.1177/002383097702000407
- Sorace, A. (2005). "Syntactic optionality at interfaces," in *Syntax and Variation: Reconciling the Biological and the Social*, eds L. Cornips and K. Corrigan (Amsterdam: John Benjamins), 46–111.
- Stein, N. L., and Glenn, C. G. (1979). "An analysis of story comprehension in elementary school children," in *Advances in Discourse Processes: New Directions in Discourse Procession*, ed. R. O. Freedle (Norwood, NJ: Ablex), 213–249.
- Strand, S., Malmberg, L., and Hall, J. (2015). *English as an Additional Language (EAL) and Educational Achievement in England: An Analysis of the National Pupil Database*. London Education Endowment Foundation.
- Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., and Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *Int. J. Lang. Commun. Disord.* 45, 480–493. doi: 10.3109/13682820903198058
- Szewczyk, J., and Wodniecka, Z. (2012) Subject- and item-related factors explaining performance in the Polish NWR test. *Paper Presented at the Sixth Working Groups and Management Committee Meeting, COST IS0806 Project*, Berlin.
- Tamburelli, M., Sanoudaki, E., Jones, G., and Sowinska, M. (2015). Acceleration in the bilingual acquisition of phonological structure: evidence from Polish–English bilingual children. *Bilingualism* 18, 713–725. doi: 10.1017/S1366728914000716
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 15, 426–445. doi: 10.1177/1367006911403202
- Thordardottir, E. (2017). Amount trumps timing in bilingual vocabulary acquisition: effects of input in simultaneous and sequential school-age bilinguals. *Int. J. Biling.* doi: 10.1177/1367006917722418 [Epub ahead of print].
- Thordardottir, E., and Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J. Commun. Disord.* 46, 1–16. doi: 10.1016/j.jcomdis.2012.08.002
- Tuller, L. (2015). "Clinical use of parental questionnaires in multilingual contexts," in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, Vol. 13, eds S. Armon-Lotem, J. de Jong and N. Meir (Bristol: Multilingual Matters), 301–330.
- Uccelli, P., and Pérez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Lang. Speech Hear. Serv. Sch.* 38, 225–236. doi: 10.1044/0161-1461(2007/024)
- Umbel, V. M., and Oller, D. (1994). Developmental changes in receptive vocabulary in Hispanic bilingual school children. *Lang. Learn.* 44, 221–242. doi: 10.1111/j.1467-1770.1994.tb01101.x
- Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: the case of Dutch gender. *Bilingualism* 16, 86–110. doi: 10.1017/S1366728912000284
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., and Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Appl. Psycholinguist.* 35, 765–805. doi: 10.1017/S0142716412000574
- van der Slik, F., van Hout, R. V., and Schepens, J. (2017). The role of morphological complexity in predicting the learnability of an additional language: the case of La (additional language) Dutch. *Second Lang. Res.* doi: 10.1177/0267658317691322
- Vender, M., Garraffa, M., Sorace, A., and Guasti, M. T. (2016). How early L2 children perform on Italian clinical markers of SLI: a study of clitic production and nonword repetition. *Clin. Linguist. Phon.* 30, 150–169. doi: 10.3109/02699206.2015.1120346
- Verhagen, J., and Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *J. Exp. Child Psychol.* 141, 65–82. doi: 10.1016/j.jecp.2015.06.015
- Verhoeven, L., Steenge, J., van Weerdenburg, M., and van Balkom, H. (2011). Assessment of second language proficiency in bilingual children with specific language impairment: a clinical perspective. *Res. Dev. Disabil.* 32, 1798–1807. doi: 10.1016/j.ridd.2011.03.010
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Appl. Psycholinguist.* 22, 217–234. doi: 10.1017/S0142716401002041
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children - Revised*. New York, NY: Psychological Corporation.
- White, A. (2011). The mobility of Polish families in the West of England: translocalism and attitudes to return. *Studia Migracyjne Przegląd Polonijny* 37, 11–32.
- Windsor, J., Kohnert, K., Lobitz, K. F., and Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *Am. J. Speech Lang. Pathol.* 19, 298–310. doi: 10.1044/1058-0360(2010/09-0064)
- Winsler, A., Díaz, R. M., Espinosa, L., and Rodriguez, J. L. (1999). When learning a second language does not mean losing the first: bilingual language development in low-income, Spanish-speaking children attending bilingual preschool. *Child Dev.* 70, 349–362. doi: 10.1111/1467-8624.t01-1-00026
- Yip, V., and Matthews, S. (2000). Syntactic transfer in a Cantonese–English bilingual child. *Bilingualism* 3, 193–208. doi: 10.1017/S136672890000033X
- Yu, B. (2013). Issues in bilingualism and heritage language maintenance: perspectives of minority-language mothers of children with autism spectrum disorders. *Am. J. Speech Lang. Pathol.* 22, 10–24. doi: 10.1044/1058-0360(2012/10-0078)

**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, MR declared a past co-authorship with two of the authors, MŁ and EH to the handling Editor.

Copyright © 2017 Haman, Wodniecka, Marecka, Szewczyk, Bialecka-Pikul, Otwinowska, Mieszkowska, Luniewska, Kolak, Mięsz, Kacprzak, Banasik and Foryś-Nogala. 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.



# Acquiring Clitic Placement in Bilectal Settings: Interactions between Social Factors

Kleanthes K. Grohmann<sup>1,2\*</sup>, Elena Papadopoulou<sup>2,3</sup> and Charalambos Themistocleous<sup>2,4</sup>

<sup>1</sup> University of Cyprus, Nicosia, Cyprus, <sup>2</sup> Cyprus Acquisition Team, Nicosia, Cyprus, <sup>3</sup> European University Cyprus, Nicosia, Cyprus, <sup>4</sup> University of Gothenburg, Gothenburg, Sweden

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University, UK

### Reviewed by:

Alexandra Perovic,  
University College London, UK  
Ur Shlonsky,  
Université de Genève,  
Switzerland

### \*Correspondence:

Kleanthes K. Grohmann  
kleanthi@ucy.ac.cy

### Specialty section:

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

**Received:** 16 December 2016

**Accepted:** 17 March 2017

**Published:** 12 April 2017

### Citation:

Grohmann KK, Papadopoulou E and  
Themistocleous C (2017)  
Acquiring Clitic Placement in  
Bilectal Settings: Interactions  
between Social Factors.  
Front. Commun. 2:5.  
doi: 10.3389/fcomm.2017.00005

This article examines the development of object clitic placement by children acquiring Cypriot Greek. Greek-speaking Cyprus is sociolinguistically characterized by diglossia between two varieties of Greek, the local Cypriot Greek and the official Standard Modern Greek. Arguably as a result of this situation, clitics may be placed postverbally (enclisis) or preverbally (proclisis) in the same syntactic environment; while the former is a property of Cypriot Greek and the latter is typically considered an effect of the standard language. The following issues are investigated here: (a) how such bilectal speakers distinguish between the two Greek varieties with respect to clitic placement; (b) how the acquisition of clitics develops over time; (c) how, and which, sociolinguistic factors determine clitic placement; and (d) how schooling may affect clitic placement. To address (a)–(d), a sentence completion task was used to elicit clitic productions, administered to 431 children around Cyprus ranging from 2 years 8 months to 8 years 11 months. The C5.0 machine-learning algorithm was employed to model the interaction of (socio-)linguistic factors on the development of clitic placement. The model shows that speakers acquire the relevant features very early, yet compartmentalization of form and function according to style emerges only as they engage in the larger speech community. In addition, the effects of sociolinguistic factors on clitic placement appear gradually.

**Keywords:** acquisition of clitics, discrete bilectalism, sociolinguistic factors in language development, C5.0 algorithm, diglossia, socio-syntax of development hypothesis

## INTRODUCTION

Language acquisition is assumed to proceed uniformly (Lenneberg, 1967). For example, across languages children between 6 and 8 months of age start to babble; at about 10–12 months, they produce and understand some words; and at around 2 years, they combine words. Even bilingual children follow the same path, though somewhat delayed [see, e.g., Tsimpli (2014) and responses]. This said, children acquiring language in bilingual settings have to tackle two major problems: (i) the extreme complexity imposed by the systemic and external variation in the input and (ii) the choice of the right code that suits the appropriate linguistic environment.

This study revisits the research presented in the study by Grohmann (2014a) on the acquisition of (preverbal vs. postverbal) object clitic placement in Cyprus, a sociolinguistically diverse environment, which is traditionally characterized by diglossia (Newton, 1972, in the sense of Ferguson, 1959), understood linguistically as “(discrete) bilectalism” (Rowe and Grohmann, 2013). In this environment, clitic placement displays features that are both Cypriot Greek (canonically enclisis) and Standard Modern Greek (proclisis by default). Unlike previous publications, this article embeds

language variation and clitic acquisition more clearly within the Socio-Syntax of Development Hypothesis (Grohmann, 2011). The previously collected data are now completely reanalyzed. In particular, the statistical methods employed here offer a model of sociosyntactic variation and constitute a new proposal for analyzing language acquisition. In this sense, this study offers new insights in the acquisition of clitics: It includes the role of gender, age, and place of origin not as isolated properties but as factors that interact dynamically and influence the acquisition and subsequent development of object clitics in Greek Cypriot children. Only in this way we can understand the Socio-Syntax of Development Hypothesis as effects of multiple and dynamic social factors on linguistic variables: Each factor can have different significance with respect to the other factors.

This article is structured as follows. After a brief presentation of bilectalism, which characterizes the linguistic landscape in Cyprus, Section “Background” provides the background on basic properties of clitic placement in the two varieties of Greek and lays out the three experimental hypotheses pursued in this study. Section “The Present Study” introduces the study, including measurements and statistics employed, followed by a presentation of the main results in Section “Results.” A thorough discussion follows in Section “Discussion,” which also sketches a sociosyntactic model for the acquisition of clitic placement. Final remarks conclude the article.

## BACKGROUND

### Discrete Bilectalism in Cyprus: Cypriot Greek (CG) and Standard Modern Greek (SMG)

Cypriot Greek is the local variety of Greek spoken in Cyprus. It is often distinguished into “village CG” and “urban CG” (Newton, 1972). However, Hadjioannou et al. (2011) suggest that post-1974, regional varieties are in the process of being leveled out due to demographic and social changes, and a Pancyprrian *koiné* variety is fast emerging [see also Tsiplakou (2014)]. They argue that urban CG—in their terminology, the Pancyprrian *koiné*, perhaps what Arvaniti (2010) calls Cypriot Standard Greek (CSG)—stands in a diglossic relationship to SMG<sup>1</sup>: CG is the sociolinguistic L(ow)-variety, and SMG is the superimposed H(igh)-variety.

Hadjioannou et al. (2011) do not do full justice to local variability, though, that exists between village CG, namely the less prestigious varieties or basilects and the most prestigious urban CG (Newton, 1972; Goutsos and Karyolemou, 2004; Arvaniti, 2010). In addition, by assuming that SMG is the acrolectal form, they arguably presuppose that CG speakers should be perfect bilinguals, which

is not the case (Arvaniti, 2010; Leivada et al., 2017, in press). Note that other than the many Greek citizens who live in Cyprus [29,321 as per the Statistical Service of Cyprus (CYSTAT) (2011)<sup>2</sup>], the Greek-speaking population of Cyprus employs the CG variety in one form or another on a regular basis, if not predominantly, in their day-to-day linguistic experiences.

Research initiated by Arvaniti (2010), including more recent work from our own research group (Grohmann and Leivada, 2012; Rowe and Grohmann, 2013, 2014; Leivada et al., 2017), provides a more refined account of the linguistic situation in Cyprus. Specifically, these studies consider as the H-variety “urban CG” or “CSG,” which is more homogeneous and gains recognition as the more prestigious form of CG, compared to village CG, which is the true L-variety. This is not a new development; the distinction between unambiguous CG forms on the one hand and unambiguous SMG forms on the other hand were not always clear-cut. For example, when Newton described the sociolinguistic situation on the island during the 1960s, he suggested that the dialect features, alongside so-called SMG features, are often blurred: “[A]part from the quite considerable gap between village dialects and the strongly standardizing speech of urban Cypriots, /xorkátika/[i.e., village CG—GPT] itself is often not heard in a pure form, but is interspersed with elements most conveniently regarded as belonging to standard Greek” (Newton, 1972, p. 108).

Furthermore, even in written language, which follows the conventions of SMG, CG speakers’ output often displays features of the dialect (Arvaniti, 2010; Leivada et al., 2017, in press). In such a complex linguistic environment, the varieties do not have well-defined boundaries. Rowe and Grohmann (2013) suggest that because Greek Cypriots eventually tease apart the varieties and render complete compartmentalizations (able to distinguish between CSG and SMG), Cyprus is (still) diglossic. To capture the linguality of Greek Cypriots, Rowe and Grohmann (2013, 2014) propose the notion of “(discrete) bilectalism.” To the extent that the term *bilectalism* is applied this way to the linguistic situation in Cyprus:

[I]t suggests dual competence of the varieties native to two polities (Greece and Cyprus) and their respective native varieties (SMG and CG). It also describes individual competencies in the two varieties [but non-randomly, that is, crucially only—GPT] as a function of these individuals living and participating in this type of society.

(Grohmann, 2014a, p. 4)

CG and SMG differ most obviously in their phonetics, (morpho)phonology, and lexicon (e.g., Newton, 1972; Theodorou, 2007; Arvaniti, 2010). As for morphosyntax, there are also a large number of differences, but it is clitic placement that has arguably drawn the greatest attention (e.g., Agouraki, 1997, 2001; Terzi,

<sup>1</sup>Ferguson (1959, p. 336) defines diglossia as “a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any section of the community for ordinary conversation.”

<sup>2</sup>Population census does not distinguish between CG and SMG; therefore, we employ citizenship as an indicative number of the population of SMG speakers. In addition, the reported number increased with the economic crisis in Greece.

1999a,b; Revithiadou, 2006; Revithiadou and Spyropoulos, 2008; Chatzikyriakidis, 2010, 2012; Pappas, 2012, 2014). The following provides a brief overview of clitic placement in the two varieties.

## Non-Uniformity of Clitic Placement in CG

Terzi (1999b) characterized CG as a Tobler–Mussafia-type language, which means that in canonical environments, clitics follow the finite verb form (enclisis) rather than precede it (proclisis). In other words, CG exhibits a pattern of mixed clitic placement, with enclisis the unmarked option and proclisis required in particular structural environments. Similar behavior in clitic placement is exhibited in European Portuguese (Duarte and Matos, 2000). In many syntactic environments that are canonical for postverbal clitic placement in CG, SMG requires preverbal clitic placement, similar to Spanish and most other Romance varieties compared to Western Romance [European Portuguese, Galician, and Asturian; cf. Lorenzo (1994)].

Modern Greek is a fairly free word order language, with SVO the most frequent and VSO another contestant for the unmarked order (e.g., Philippaki-Warbuton, 1985; Lascaratos, 1998; Roussou and Tsimpli, 2006); no conclusive evidence has been presented on the latter issue or on possible differences between the two varieties (but see, among others, the study by Terzi (1999a) who argues for a different landing site of the finite verb). In CG, as in SMG, third-person direct object clitics are derived from strong pronouns; clitics are marked for number, gender, and case. Concerning the particular characteristics of mixed clitic placement, it can be observed that certain syntactic environments enforce preverbal placement—otherwise enclisis is found. Therefore, clitics in CG can appear postverbally in both imperative and non-imperative contexts, whereas in SMG, they can appear only as enclitics in imperatives and gerunds.

The data below, taken from the study by Theodorou and Grohmann (2015), illustrate the relevant differences between the two varieties across different syntactic environments, starting with a declarative context in indicative mood.

|                                |                                                                                                        |          |
|--------------------------------|--------------------------------------------------------------------------------------------------------|----------|
| (1) <i>Indicative context:</i> |                                                                                                        |          |
| a.                             | (O Jannis) <i>θcavazi/δjavazi to vivlio.</i><br>the John reads the book<br>“John is reading the book.” | [CG/SMG] |
| b.                             | (O Jannis) <i>θcavazi to.</i><br>the John reads it-CL<br>“John is reading it.”                         | [CG]     |
| c.                             | (O Jannis) <i>to δjavazi.</i><br>the John it-CL reads<br>“John is reading it.”                         | [SMG]    |

In some contexts, enclisis is the only grammatical option in both varieties:

|                                |                                                    |       |
|--------------------------------|----------------------------------------------------|-------|
| (2) <i>Imperative context:</i> |                                                    |       |
| a.                             | <i>θcavase to!</i>                                 | [CG]  |
| b.                             | <i>δjavase to!</i><br>read.IMP it-CL<br>“Read it!” | [SMG] |

|                               |                                                             |       |
|-------------------------------|-------------------------------------------------------------|-------|
| (3) <i>Gerundive context:</i> |                                                             |       |
| a.                            | <i>θcavazontas to</i>                                       | [CG]  |
| b.                            | <i>δjavazontas to</i><br>reading.PART it-CL<br>“reading it” | [SMG] |

In others, clitics appear preverbally even in CG, namely when a linguistic expression appears in the left periphery of the clause—in particular, *wh*-elements and relative operators trigger proclisis in CG; the same holds for negative contexts and the subjunctive marker *na*. This is exemplified for both Greek varieties in (4), where we only signal the different phonetic forms of the verb (and negation).

|                              |                                                                                                                                |          |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------|----------|
| (4) a. <i>wh-question</i>    |                                                                                                                                |          |
|                              | Pu <i>to</i> <i>θcavazi/δjavazi</i> (o Jannis)?<br>where it-CL reads (the John)?<br>“Where does John read it?”                 | [CG/SMG] |
| b. <i>Relative clause</i>    |                                                                                                                                |          |
|                              | o andras pu <i>to</i> <i>θcavazi/δjavazi</i><br>the man that it-CL reads<br>“the man who/that reads it”                        | [CG/SMG] |
| c. <i>Negative clause</i>    |                                                                                                                                |          |
|                              | En/ōen <i>to</i> <i>θcavazi/δjavazi</i> (o Jannis).<br>not it-CL reads (the John)<br>“Jannis doesn’t read it.”                 | [CG/SMG] |
| d. <i>Subjunctive clause</i> |                                                                                                                                |          |
|                              | Perimeno na <i>to</i> <i>θcavasi/δjavasi</i> (o Jannis).<br>expect to it-CL read (the John)<br>“I expect [Jannis to read it].” | [CG/SMG] |

This brief exposition suffices for present purposes, since the only environment tested in the present experimental design is a version of (1), namely a declarative context in indicative mood. Focus of this article is the acquisition and subsequent development of clitic production by children acquiring CG. In that context, a closer examination of clitic placement reveals three notable properties. The first is that, while canonically enclitic, CG requires proclisis in certain syntactic environments; children thus have to master the different environments that trigger CG placement positions. The second property is that in most environments, CG clitic placement and SMG clitic placement are identical; clear examples are provided by enclisis portrayed in (2) and (3) and proclisis in (4a–d). Third, there are environments that trigger enclisis in CG but proclisis in SMG, which, consequently, constitute a potential source of variability in speakers’ speech; this is the canonical case of declaratives in indicative mood such as in (1).

What previous research from our research group has shown is that CG-speaking children very often mix clitic placement. That is, the same child would respond with enclisis in one case and proclisis in another. The main finding of Grohmann’s (2011) pilot study was that, while 3 and 4 year olds as well as adults consistently employed enclisis, 5-year-old children fell into three groups: roughly 40% consistently employed enclisis, around 40% consistently employed proclisis, and the remaining 20% mixed the two to a large extent. This general pattern depicted in **Figure 1** was confirmed by subsequent testing, summarized by Grohmann (2014a) who also compares the mean numbers of (non-target) proclisis across our different studies, presented in **Table 1**.

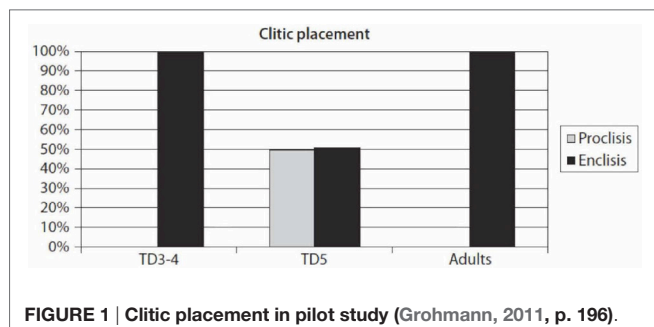


**TABLE 1 | Non-target placement across studies (Grohmann, 2014a, 25).**

| CAT study                                                      | Mean age in months (SD) | Mean preverbal clitic placement out of 12 (SD) |
|----------------------------------------------------------------|-------------------------|------------------------------------------------|
| Grohmann et al. (2012)                                         | 60.83 (13.56)           | 3.28 (4.08)                                    |
| Agathocleous (2012)                                            | 66.53 (10.25)           | 4.08 (4.73)                                    |
| Leivada et al. (2010)                                          | 73.18 (19.61)           | 4.98 (5.02)                                    |
| Charalambous and Agathocleous (2011, Unpublished) <sup>1</sup> | 79.29 (10.95)           | 6.00 (5.08)                                    |
| Theodorou (2013)                                               | 84.54 (15.54)           | 7.73 (4.67)                                    |
| Charalambous and Agathocleous (2012, Unpublished) <sup>2</sup> | 86.76 (9.00)            | 6.68 (5.01)                                    |

<sup>1</sup>Charalambous, A. and M. Agathocleous (2011). *The acquisition of clitics in Cypriot Greek children's language living in a rural setting*. Unpublished ms., University of Cyprus, Nicosia.

<sup>2</sup>Charalambous, A. and M. Agathocleous (2012). *The development and the role of the social environment in Cypriot Greek clitic placement: Factors and trends*. Unpublished ms., University of Cyprus, Nicosia.

Apparent optionality in clitic placement in certain syntactic environments has also occasionally been noted for adult CG, culminating in separate (sociolinguistic) empirical investigations in which Pappas (2012, 2014) reports a certain level of variability. Yet, it would be misleading to characterize the two options of clitic placement as syntactic variants; after all, our research team's production tasks administered in CG almost invariably prompted enclisis, whereas the same task carried out in SMG led to proclisis (Leivada et al., 2010). This suggests that CG and SMG are discretely distinguished by speakers, each with its own grammatical rules. To the extent that there may be a blur between proclisis and enclisis in the speech of Greek Cypriots, we contend that this must be due to the still poorly understood exact acrolect or H-variety in the guise of CSG, SMG, or some mixed form(s). Future research across generations of speakers needs to clarify this issue employing insights from sociolinguistics and theoretical linguistics.

As Leivada and Grohmann (in press) note, there might be an obvious way to approach the situation in which both enclisis and proclisis are encountered in identical syntactic contexts by the same speakers: One might appeal to "competing grammars," a concept going back to Kroch (1994), who proposed competition of grammatical systems in diachronic change "between grammatically incompatible options which substitute for one another in usage" (p. 180); for specific accounts and extensions to language acquisition models, see Kroch and Taylor (2000),

Yang (2000), Legate and Yang (2007), among others. Note that Lightfoot (1999) characterized competing grammars to reflect "internalized diglossia"; hence, this might indeed be an appropriate approach to take up for CG. Competition of the CG and SMG grammatical systems has been explicitly suggested by Tsiplakou (2009, 2014). In our own work, we further enrich the model by integrating two related older notions, "competing motivations" (Du Bois, 1985) and "metalinguistic awareness" (Cazden, 1976).

Moreover, the competing grammars hypothesis would have children acquiring the native CG grammar (enclisis) face the emerging SMG grammar (proclisis). This just happens to grow stronger through increased input. Since formal schooling is carried out, by law, in the medium of SMG (but see Sophocleous, 2011), it is around the entrance into the school system that the SMG grammar becomes stronger, perhaps even dominant at times. We will turn to this next.

## Toward Capturing the Socio-Syntax of Development Hypothesis

In an at first glance alternative approach, Grohmann (2011) proposed the Socio-Syntax of Development Hypothesis. According to subsequent refinements,

the [Socio-Syntax of Development Hypothesis] approaches the acquisition of syntactic variants that pertain to different varieties, in bi-*x* environments, as proceeding through the existence of competing motivations that arise depending on the level of proximity (in the dialectal continuum) existing between the variety that the child is exposed to prior entering school and the one used in school—that is, even beyond the 'normal' period of native language acquisition.

(Grohmann and Leivada, 2012, p. 257)

In essence, the Socio-Syntax of Development Hypothesis holds that in the course of the language acquisition process, emergence of sociolinguistic and metalinguistic awareness as well as development of social identities account for children's sociolinguistic choices. Because both metalinguistic awareness and language acquisition develop over time, we expect this to be reflected in children's different linguistic choices. Moreover, we expect that, to a certain degree, social factors account for the particular choices (Nesdale and Rooney, 1996; Habib, 2016). An experiment has been designed to test these expectations. In the next section, we discuss the experimental hypotheses of this study.

Many sociolinguistic studies have investigated, among others, sex (male/female) or, as used here, gender differences between boys and girls in language development across heterogeneous populations. Earlier studies suggest that gender is a fundamental factor for variation (Maccoby and Jacklin, 1974; Fenson et al., 1994), whereas more recent ones claim the exact opposite (Hyde, 2005; Wallentin, 2009). In an attempt to resolve the debate, Barbu et al. (2015) investigated not only gender differences but also children's socioeconomic status (SES) who acquired a frequent phonological alternation in French between the ages of around 2.5 and 6.5 years. Gender differences were only found for children with low SES, whereas low-SES boys performed worse

than low-SES girls. No such differences were observed in children with higher SES, suggesting the need for a better reorganization of conditions tested in language development.

Another factor commonly assumed to account for linguistic variation is the urban/rural dichotomy of children's habitation. A recent study by Habib (2016) showed the continuous use of urban Arabic features in rural children. In particular, girls predominantly used their mothers' urban feature, the glottal stop [ʔ], in place of the rural voiceless uvular stop [q]. In contrast, boys reverted to the use of the rural variant. The distribution of these variants does not only depend on gender: The younger the children were, the more urban variants they produced, displaying a great decrease of urban variant production in the group of 12 to 14 year olds whose rural variant production was increased instead.

A further important aspect of sociolinguistic and educational linguistic research concerns the role of literacy in dialectal contexts. A pioneer in the area within the United States, Wolfram (1994) investigated the relationship between bidialectalism and literacy, mostly in pupil populations from African American Vernacular English-speaking background. Linguistic and cultural differences were found to be factors in poor reading abilities, but the work also offers perspectives on language variation for practitioners through noting grammatical differences in dialects or implications for instruction and assessment.

In Europe, too, the effect of monodialectal vs. bidialectal literacy is now being explored. To mention just one study, Vangsnes et al. (2017) deal with the two written standards of Norwegian, Bokmål (majority variety), and Nynorsk (minority variety). Children in Norway are schooled in one or the other variety, yet pupils schooled in Nynorsk acquire Bokmål simultaneously through extracurricular exposure, hence develop "bidialectal literacy." The authors correlate the results from standardized national tests in reading, arithmetic, and English for eighth graders, with information available on language of instruction and SES. The main finding is that Nynorsk pupils perform better than average, which the authors take to be an effect of the "bilingual advantage" in cognitive development—and, importantly with relevance for this study, that such an advantage may arise even in the case of two closely related varieties.

For further interesting research on the relation between executive control and language abilities in two closely related varieties, Sardinian and Italian, in the context of schooling, see the study by Garraffa et al. (2015). Although not directly linked to executive control abilities (i.e., the "bilingual advantage" in cognitive development), there is a growing body of work on literary development in Cyprus, too, which is sensitive to the native CG variety in the context of SMG-dominant reading and writing instruction in school (e.g., Tsiplakou, 2006; Hadjioannou et al., 2011). Current research from Greece for SMG connects performance on executive control tasks explicitly with literary skills for monolingual and bilingual children (Andreou, 2015; Andreou and Tsimpli, unpublished<sup>3</sup>). Following up on a first study by Antoniou et al. (2016), this connection is also being investigated for bilectal,

bilingual, and monolingual children in several ongoing dissertations within the Cyprus Acquisition Team (CAT Lab).

In addition, the issue of schooling as a factor on language development does seem to get some recognition lately, even beyond the area of literacy and in more formal approaches to language acquisition. For example, Heycock et al. (2013) look at (new) language change on the Faroe islands, where Danish seems to influence the language of Faroese-speaking children with the onset of schooling—that is, as in Cyprus, clearly after the critical period of the first-language acquisition process. They tested 5- to 10-year-old Faroese children on grammaticality judgments and elicited productions of subordinate clauses, which seem to be undergoing a change from an Icelandic-like system to one like Danish, a change away from V to T. The study shows that preschool children exhibit more of the Icelandic-like order than adults do, that is, the change takes place later [see also the study by Heycock and Sorace (unpublished)<sup>4</sup> on embedded V2 across the Scandinavian languages, where judgments tend to be gradient rather than variable].

The starting point of efforts exploring the "cognitive advantage" of bilectal children from Cyprus reported in the study by Antoniou et al. (2016) comes from the original findings on clitic placement by young school-age children (Grohmann, 2011), which has subsequently been researched with many more child populations and groups [summarized in the study by Grohmann (2014a)]. This article constitutes a comprehensive overview of the reported data with a novel statistical approach. Thus, it lays the foundations for the current research agenda of "comparative bilingualism" (Grohmann, 2014b) and the gradience of multilingualism across populations (Grohmann and Kambanaros, 2015), even if that is in and of itself not the focus of this contribution. As introduced above, the present focus is the acquisition and subsequent development of clitic production by children acquiring CG. Our aims are to first address the issue of how clitic placement interacts with factors such as *gender*, *place of residence*, and *input factors*, which govern language acquisition, and then to provide a model that predicts the development of clitic placement in CG.

## Experimental Hypotheses

Before the age of 3 years, both CG- and SMG-acquiring children have mastered clitic production (Marinis, 2000; Petinou and Terzi, 2002; Leivada et al., 2010; Grohmann, 2011; Grohmann et al., 2012). Given the variation in a child's (socio)linguistic environment, a theory of language acquisition must thus provide an account of the cognitive processes, which take place when children employ alternative structures in their speech, and an account of the child's form-decision strategies in bilectal (bidialectal, bivariational, and possibly bilingual) settings that allow for alternative forms to coexist.

Our aim is twofold: to first address the puzzling issue of how clitic placement, a grammatical characteristic, interacts with factors such as *gender*, *place of residence*, and *input factors* that

<sup>3</sup>Andreou, M., and Tsimpli, I. M. (2015). Dominance, biliteracy and cognitive control: effects on bilingual children's narratives. Unpublished manuscript.

<sup>4</sup>Heycock, C., and Sorace, A. (2007). Verb movement in Faroese: new perspectives on an old question. Unpublished manuscript.

govern language acquisition and then to provide a model that predicts the development of clitic placement in CG. We will argue that the linguistic settings constitute competitive sociolinguistic environments in which linguistic codes and forms are in conflict and that these environments give rise to (sociolinguistic) decisions and learning solutions.

In line with these goals, the following three experimental hypotheses will be pursued to account for clitic placement in CG:

- (H1) *Hypothesis 1: Proclisis increases with chronological age.* In fact, Grohmann et al. (2012) and Agathocleous et al. (2014) demonstrated that, as they grow up, children acquiring CG employ proclisis more often; this tendency culminates in teenagers who show 100% preference in enclisis. If this hypothesis is correct, chronological age is expected to determine clitic placement for two reasons: (i) because children master clitic placement as a part of their linguistic development and (ii) because, as they grow up, children acquire the sociolinguistic norms of their community. Therefore, linguistic properties that associate with the H-variety, such as proclisis, will steadily increase in formal settings (including psycholinguistic experiments), whereas features that are associated with the L-variety, such as enclisis, will decrease.
- (H2) *Hypothesis 2: Proclisis increases with schooling level.* This hypothesis is in line with the Socio-Syntax of Development Hypothesis (Grohmann, 2011). Because SMG (or CSG with SMG-like rules that differ from non-acrolectal CG in the relevant respect) is the standard variety and taught in schools, we expect that systematic teaching of SMG would effectuate the use of proclisis; otherwise proclisis will remain constant or will be used at a lesser degree than enclisis.
- (H3) *Hypothesis 3: The degree of proclisis and enclisis use depends on sociolinguistic factors.* Accordingly, we expect the following:
- (i) Because enclisis is a property of an L-variety, it is expected to index *masculinity*, whereas proclisis, associated with the H-variety, is expected to index *femininity*. This claim is based on observations that women employ more standard variants than men (Trudgill, 1972; Labov, 2001), especially if the standard is also an innovative form for a given language community. Accordingly, we hypothesize that boys will employ enclisis as a male variant, whereas girls will employ proclisis as a female variant.
  - (ii) Speakers from rural areas will employ more enclisis to a greater degree than speakers who live in urban areas. This claim is in line with studies that suggest a distinction between village CG and urban CG (e.g., Newton, 1972; Hadjioannou et al., 2011); previous research on CG clitic acquisition suggests that *urban* vs. *rural place of residence* influences clitic placement (Agathocleous et al., 2014).
  - (iii) Gender and place of residence are not simplex phenomena but interact with each other (Eckert, 1999); hence, it is the *interaction of these sociolinguistic factors* that accounts for the acquisition of clitic placement.

## THE PRESENT STUDY

### Participants

In the testing period from 2008 to 2011, a total of 431 children participated in the experiment; for a detailed description of all populations tested on this experiment and the CAT Clitics Corpus and references to published analyses, see the study by Grohmann (2014a). At the time of data collection, all participants were aged between 2 years and 8 months and 8 years and 11 months; they were all native acquirers of Cypriot Greek, had two Greek Cypriot parents, and were born and raised in CG-speaking Cyprus. The research was approved by the Cyprus Ministry of Education and Culture upon submission of the full description of the tool and protocol to the Pedagogical Institute. Parents and participating schools' headmasters and teachers involved provided their written consent after detailed letters of information concerning the research to be conducted; hence, additional ethics approval was deemed unnecessary at the time of data collection. **Table 2** shows the distribution of speakers across different age groups, arranged here by chronological age in years.

To elicit responses from all CG varieties, data collection took place in both urban and rural areas across (the Greek-speaking part of) the island. **Table 3** shows the number of speakers who participated in the study based on their place of residence (in alphabetical order).

### Methodology

This study adapted the COST Action A33 Clitics-in-Islands testing tool (Varlokosta et al., 2016) so as to elicit the production of third person singular direct object clitics.<sup>5</sup> The task comprises 19 items, 12 target structures preceded by two warm-ups and interspersed by 5 fillers. All target and warm-up structures were declarative sentences with a transitive verb, with one half in present tense and the other in past tense, as in (5) and its corresponding picture in **Figure 2**.

- (5) I korua luni tin kamiloparðali tfe i kamiloparðali engathari. Jati i kamilo parðali engathari? I kamiloparðali engathari jati i korua... [target response: *plinisci tin or luni tin*]

"The girl is washing the giraffe and the giraffe is clean. Why is the giraffe clean? The giraffe is clean because the girl..." [target response: *washes it (= her-CL)*]

The target clitic pronoun was produced inside a *because*-clause and invariably referred to a third person singular object mentioned in the experimenter's introduction.

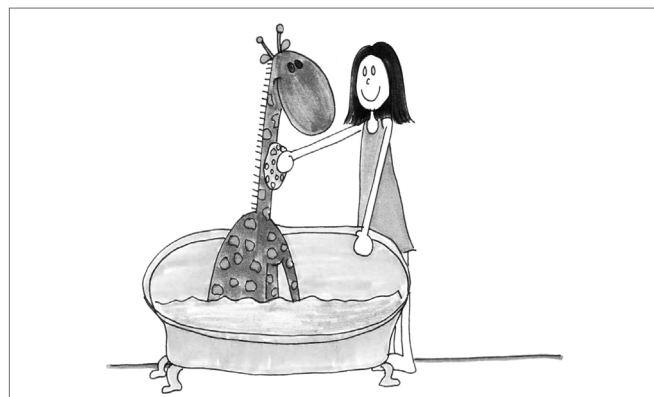
<sup>5</sup>For the development of the tool within COST Action A33, with a clear cross-linguistic intention, the *because*-island was chosen to provide a context for obligatory clitic use (Varlokosta et al., 2016). The intention of setting up a syntactic island environment was to elicit clitics even in languages that frequently allow object drop, including the grammatical omission of clitics in European Portuguese, where it is supposed to be ungrammatical (Raposo, 1986). For discussion, see the study by Costa and Lobo (2014) who argue that comprehension tasks are better suited to detect the mastery of clitic production and placement, whether in regular object drop or in more complex island contexts.

**TABLE 2 | Participants.**

| Participants | Number | Age range (mean)                                  | Age range (SD) | Gender [female (F)/male (M)] | Urban/rural              |
|--------------|--------|---------------------------------------------------|----------------|------------------------------|--------------------------|
| 2-year olds  | 18     | 2 years 8 months to 2 years 11 months (38 months) | 3.5 months     | 9F<br>9M                     | 7U<br>2R<br>8U<br>1R     |
| 3-year olds  | 58     | 3 years to 3 years 11 months (51 months)          | 3.1 months     | 28F<br>30M                   | 14U<br>14R<br>17U<br>13R |
| 4-year olds  | 118    | 4 years to 4 years 11 months (62 months)          | 3.5 months     | 68F<br>50M                   | 37U<br>31R<br>29U<br>21R |
| 5-year olds  | 120    | 5 years to 5 years 11 months (73 months)          | 3.8 months     | 60F<br>60M                   | 44U<br>16R<br>40U<br>20R |
| 6-year olds  | 75     | 6 years to 6 years 11 months (83 months)          | 3.1 months     | 40F<br>35M                   | 34U<br>6R<br>19U<br>16R  |
| 7-year olds  | 32     | 7 years to 7 years 11 months (96 months)          | 3.9 months     | 15F<br>17M                   | 10U<br>5R<br>14U<br>3R   |
| 8-year olds  | 10     | 8 years to 8 years 11 months (106 months)         | 1.5 months     | 7F<br>3M                     | 5U<br>2R<br>2U<br>1R     |
| Total        | 431    |                                                   |                | 227F<br>204M                 | 280U<br>151R             |

**TABLE 3 | Number of speakers by gender × place of residence.**

| Place (population) | Female | Male |
|--------------------|--------|------|
| Agia Fyla (14,451) | 6      | 4    |
| Evrixu (827)       | 3      | 3    |
| Ipsonas (11,117)   | 2      | 2    |
| Kolossi (5,651)    | 9      | 9    |
| Larnaka (51,468)   | 37     | 40   |
| Limassol (101,000) | 44     | 42   |
| Nicosia (55,014)   | 53     | 39   |
| Paphos (32,892)    | 33     | 22   |
| Paralimni (14,963) | 29     | 30   |

**FIGURE 2 | Test picture used from COST A33 clitics-in-islands testing tool.**

All tests were carried out by native speakers of CG. A total of five undergraduate and postgraduate students collected the data reported here. Testing lasted no longer than 5 min and was conducted in one session in a quiet room individually (child and researcher). Most children were tested at school, but a few younger ones were tested in their homes. To avoid a formal setting as much as possible and to obtain some kind of familiarity between experimenter and child, a brief conversation about a familiar topic, such as the child's favorite cartoons, took place before the testing.

The experimenter described each picture and then asked the participants to fill in the *because*-clause [see (5)]. The use of a clitic was expected; nevertheless, children also provided other responses: Some repeated *jati* "because" on their own, others filled in right after the experimenter's prompt of *jati*, and yet other children completed the sentence after the experimenter continued with the subject. In some instances, mainly after the third test item, children produced the clitic followed or preceded by the verb right after the question asked by the experimenter (*Why is the giraffe clean?*), that is, before the experimenter started uttering the *because*-clause. See also the study by Varlokosta et al. (2016) for more details.

No verbal reinforcement was provided during the test items other than encouragement with head nods and fillers. Participants received verbal feedback only during the two warm-ups at the beginning of the session. Self-correction was not registered; only the first response was recorded and used for data collection and analysis purposes. During the session, the researcher recorded the answers on a score sheet by hand; there was no audio or video recording.

## Measurements

The total number of participants amounts to 431 children; yet for the analysis, a number of productions were excluded based on the following criteria:

- The two warm-up items were removed from the data set, resulting in 12 test items.
- Productions that did not include a clitic were excluded from the data set.
- Productions with clitic placement restricted to a different environment were ignored (imperatives, subjunctives, and negation).

In total, 1,420 observations were excluded: 862 warm-up items, 429 productions that did not include a clitic, and 129 productions with clitic placement in a different environment. Thus, the final set comprises 5,580 observations. A database with the children's responses along with other sociolinguistic information was created. The database includes metadata such as the following:

- A. Participants' response (12 levels/items); 12 target structures.
- B. Date of testing; this provides the date of each testing session.
- C. Researcher; to control for the effects of the researcher on clitic placements, code names for the five researchers who ran the experiment were encoded in the database.

In addition, the database contains the predictors and the response variables. The predictors were the following:

|     |                                   |
|-----|-----------------------------------|
| (6) | <i>Speaker-Related Predictors</i> |
| a.  | Gender (two levels: Female/Male)  |
| b.  | Date of Birth and Age (in Months) |
| c.  | Age Group (5-month groups)        |

We employed a 5-month cut to facilitate more homogeneous grouping of children across age groups. This cut allowed for a more precise description of the groups and captured the gradual change in children's linguistic behavior, which is a result of the rapid linguistic development observed in children between 2 and 5 years of age.

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (7) | <i>Demographic and geographical predictors</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| a.  | School Grade (6 levels: Nursery/Kindergarten/Pre-Primary/Grade 1/Grade 2/Grade 3). School grades associate with age. Observe that school grades presupposed different learning levels; therefore, children are expected to develop preverbal clitic placement as part of an SMG-learning process as they attend higher grades (see Grohmann, 2011). School would then boost the use of preverbal placement. In other words, we expect that school will embolden the learning of clitic placement in line with the Socio-Syntax of Development Hypothesis (Grohmann, 2011; Grohmann et al., 2012). |
| b.  | Residence (two levels: Town/Village). CG is often distinguished into rural/village varieties and urban/town CG (Newton, 1972) or CSG (Arvaniti, 2010). The latter is the acrolectal form of the variety, whereas rural/village varieties are considered more basilectal. Speakers who live in rural areas are expected to employ enclisis more frequently than speakers who reside in urban centers (Agathocleous et al., 2014).                                                                                                                                                                  |
| c.  | Area (2 levels: Rural/Urban). Because not all villages are considered rural areas, this factor aims to provide a better account for speakers' regional accents and registers.                                                                                                                                                                                                                                                                                                                                                                                                                     |

The response variables are the following:

|     |                                                                                                                             |
|-----|-----------------------------------------------------------------------------------------------------------------------------|
| (8) | a. Response (the actual response provided by children).                                                                     |
|     | b. Clitic Production (3 levels: Production of clitic/No production of clitic/Production of clitic in an obligatory context) |
|     | c. Clitic placement (2 levels: Preverbal placement/Postverbal placement)                                                    |

## Statistics

To estimate the contribution of the predictors such as *School Grade*, *Gender*, *Residence*, and *Age Group* on the classification of *Clitic Placement*, we employed the machine learning and classification algorithm C5.0, the output of which is a *decision tree* (Russell and Norvig, 2003, p. 653–677). For estimating the accuracy of the model, 90% of the data was run as a training test and the remaining 10% as a test set. To provide greater accuracy and better weighting, the model was enhanced using *boosting* with 100 trials, which results in more than one decision tree (Cohen, 1995). Note that C5.0 employs winnowing, which removes all those attributes that may be unhelpful. The final attributes employed in the model are thus those that contribute to the classification. The statistical analysis was implemented in R (R Core Team, 2012), with the R package C50 for the classification.

## RESULTS

### Speaker-Related Predictors

#### Age

The findings from our earlier work (Grohmann, 2011) could be corroborated for the larger data set: Postverbal clitics clearly outnumber preverbal clitics in the responses of children aged between 30 and 40 months (**Figure 3**). Nevertheless, in the responses starting at 40 months, children's postverbal clitics steadily decrease, before becoming constant only between 45 and 50 months. Importantly, as the children grow up, the frequency of preverbal clitics increases in the responses, whereas postverbal clitics decrease.

Most importantly, **Figure 3** shows a gradual decrease of postverbal alongside the corresponding rise of preverbal clitic placement. This transition takes place between 75 and 90 months of age. After 90 months (7 years 6 months), a dramatic decrease of postverbal placement takes place. After 100 months (8 years 6 months), postverbal placement increases again. We expect this rise to continue until 100% postverbal production achieved during puberty, as observed in the studies by Grohmann et al. (2012) and Agathocleous et al. (2014), who provide data, using the same testing tool, from adolescents who clearly settle for CG enclisis only.

#### Gender

The results concerning the effect of gender on clitic placement are reported in **Table 4**. The percentage of postverbal clitic placement employed by male speakers (64%) is greater than the percentage of postverbal clitic placement employed by female speakers (58%). Most importantly, clitic placement differs significantly by gender,  $\chi^2(1) = 22.02$ ,  $p < 0.001$ .

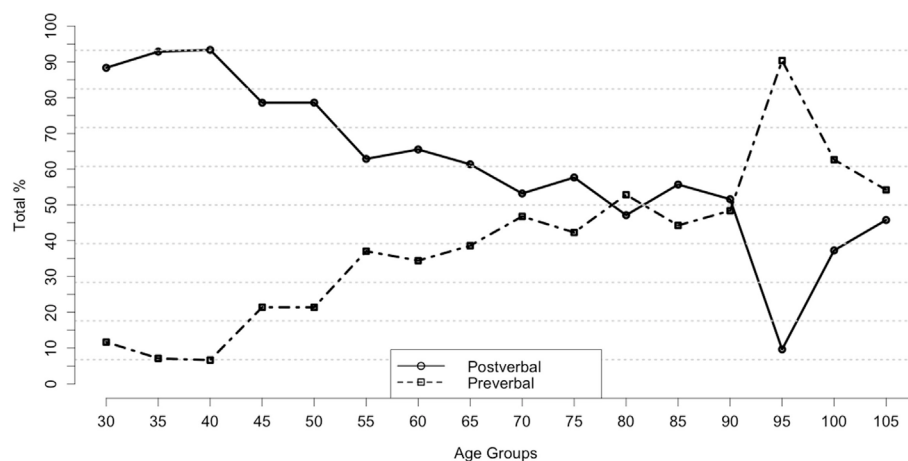


FIGURE 3 | Postverbal vs. preverbal clitic placement across 5-month age groups.

TABLE 4 | Clitic placement and gender.

| Placement  | Female |    | Male  |    | Total |
|------------|--------|----|-------|----|-------|
|            | N      | %  | N     | %  |       |
| Postverbal | 1,705  | 58 | 1,665 | 64 | 3,370 |
| Preverbal  | 1,260  | 42 | 950   | 36 | 2,210 |

## Demographic and Geographical Predictors

### Schooling

Postverbal clitics outnumber preverbal clitics in the responses of children who attend nursery school (Figure 4). At kindergarten, caregivers rather than schools influence children's speech. Nevertheless, after kindergarten, preverbal clitics gradually appear more frequently in their speech, and by second grade, preverbal clitics are outnumbered by postverbal clitics (see also Grohmann et al., 2012; Agathocleous et al., 2014).

### Place of Residence

Overall, speakers who live in towns employ fewer postverbal clitics (58.5%) than speakers who live in villages (64%). A Pearson chi-square test shows that the effect of place of residence on clitic placement is highly significant,  $\chi^2(1) = 14.69$ ,  $p < 0.001$ .

Table 5 shows the effect of place of residence on clitic placement. This factor distinguishes speakers who reside in villages with rural accents from those who live in urban centers, and it better depicts regional sociolinguistic variation than the Town–Village dichotomy presented above. A Pearson chi-square test shows that the effect of place of residence on clitic placement is highly significant,  $\chi^2(1) = 13.85$ ,  $p < 0.001$ .

## Decision Trees and Classification Structures

The classification was performed with C5.0 (accuracy = 0.7, 95% CI 0.6348–0.7537, kappa = 0.4). The boosting for the classification was reduced to seven trials, as the other trials had no contribution. The attribute usage was the 100.00% for *School Grade*,

100.00% for *Chronological Age*, 86.74% for *Residence*, and 85.49% for *Gender*.

In line with the preceding discussion, the decision tree shows that children younger than 50 months employ postverbal clitics, clearly conforming to CG (Figure 5).

Children of 50 months and older show greater variation in their choices. Specifically, the children who reside in villages and are younger than 90 months employ postverbal clitics, whereas children older than 90 months employ preverbal clitics. Gender becomes relevant for the selection of clitic placement in children who are born in towns. Boys attending first grade primarily employ postverbal clitics, whereas boys who attend nursery, kindergarten, pre-primary school, or primary school grades 1 and 2 primarily employ preverbal clitics. Girls younger than 60 months employ primarily postverbal clitics, whereas girls older than 65 months employ primarily preverbal clitics (i.e., possibly closer to SMG).

## DISCUSSION

A theory of grammar should, on the one hand, explain children's ability to acquire their native language along with its core features and parametric intricacies and, on the other, account for their unique capacity to apply alternative forms in the appropriate contexts. This study shows that children learn the appropriate social settings very early in their lives and employ the suitable choices in the two alternative clitic placements. We suggest that the Socio-Syntax of Development Hypothesis achieves considerable explanatory success, corroborated by the findings of this study.

Returning to the issue of competing grammars, we suggest that the Socio-Syntax of Development Hypothesis constitutes an explicit “trigger” for the competition between two closely related grammars. In the present context, by identifying a schooling factor in the development of CG-speaking children's grammar, we can pinpoint the time frame in which the two systems (CG and SMG) compete and why so. Note that this grammatical development takes place past the critical period and arguably does so

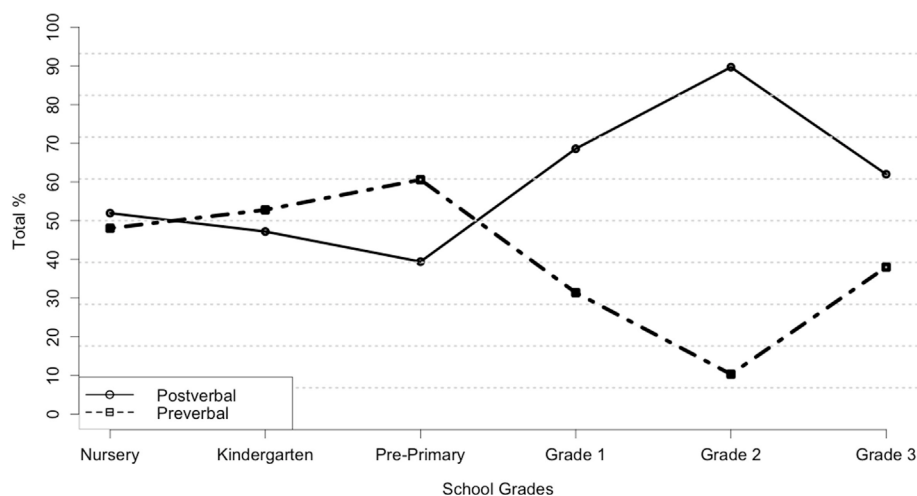


FIGURE 4 | Postverbal vs. preverbal clitic placement across school grades.

TABLE 5 | Clitic placement and place of residence (town or village).

| Placement  | Town  |      | Village |    | Total |
|------------|-------|------|---------|----|-------|
|            | N     | %    | N       | %  |       |
| Postverbal | 2,127 | 58.5 | 1,243   | 64 | 3,370 |
| Preverbal  | 1,509 | 41.5 | 701     | 36 | 2,210 |
| Total      | 3,636 |      | 1,944   |    | 5,580 |

in conjunction with “competing motivations” (Grohmann and Leivada, 2012; Leivada and Grohmann, in press). These presumably stem from the (at least) two grammars in the bilectal child’s linguistic development that compete with each other.

By showing that the linguistic choices of children depend on social factors, the tree-based model corroborates Grohmann’s (Grohmann, 2011) and Grohmann and Leivada’s (Grohmann and Leivada, 2012) formulations of the Socio-Syntax of Development Hypothesis by highlighting the interdependence between social factors and the acquisition of syntax. As such, it also demonstrates speakers’ conditional adaptation to the microsociolinguistic environment. The microsociolinguistic environment depends on stable sociolinguistic environments that affect language learner linguistic habits during acquisition process. These environments include families, sociolinguistic communities, and periods of dramatic changes such as change to a new place (i.e., school), accompanied by a change in roles, sociolinguistic identities, power relationships, and so on, which call for adaptation in the child’s sociolinguistic behavior. The conditional adaptation is what determines the choice of language form and triggers a dynamic break of a gradual process. To achieve this conditional adaptation, the speaker–hearer has to construct a representation of the environment and to employ this representation to assess the output productions. We suggest that this adaptation involves a learning procedure that accounts for parameter setting and also takes into account social variation to assess the use of the output forms. That is, we propose that the Socio-Syntax of Development Hypothesis accounts for the acquisition of linguistic phenomena

that depend on which competing grammars will surface in a specific sociolinguistic context.

Next, we examine what the results of this study tell us about the experimental hypotheses put forth in Section “Experimental Hypotheses” above and the consequences of the model of these interactions. First, let us consider (H1). The results demonstrate that chronological age affects clitic placement acquisition, thus corroborating *Hypothesis 1* (Grohmann et al., 2012; Agathocleous et al., 2014). Arguably, since clitic placement depends on the type of data (H-variety, L-variety, etc.), this effect is not simply quantitative but qualitative. Indeed, the results demonstrate that, as children grow up, proclisis associated with the H-variety (i.e., SMG or something close to it such as CSG) steadily increases in children’s speech, whereas enclisis associated with the L-variety (i.e., CG) decreases. These observations suggest that children not only acquire the social norms of their communities along with their physiological, cognitive, and linguistic maturation but also observe these social norms in their speech.

What is more, the findings corroborate (H2), namely that proclisis increases as children enter primary school (Grohmann et al., 2012; Agathocleous et al., 2014; Grohmann, 2014a). Children attending nursery school employ primarily postverbal patterns, which adhere to CG grammar, whereas preverbal placement appears to a lesser degree. After nursery school, as children attend higher grades, a gradual increase of proclisis and a gradual decrease of enclisis take place. At grade 1, children start to employ more preverbal than postverbal clitics in the present experimental context. This tendency continues in grades 2 and 3.

If we assume that proclisis is a property of SMG, its acquisition seems to depend on the role of formal education in the bilectal setting of Cyprus. Schools teach the standard language of the speech community, SMG (but see the study by Sophocleous (2011) and related studies). The role of formal education is considered an important contributing factor for L2 acquisition. Another issue that concerned especially studies on second dialect acquisition is whether the use of the native language or variety effectuates the acquisition of a second language or variety (Siegel, 2010).

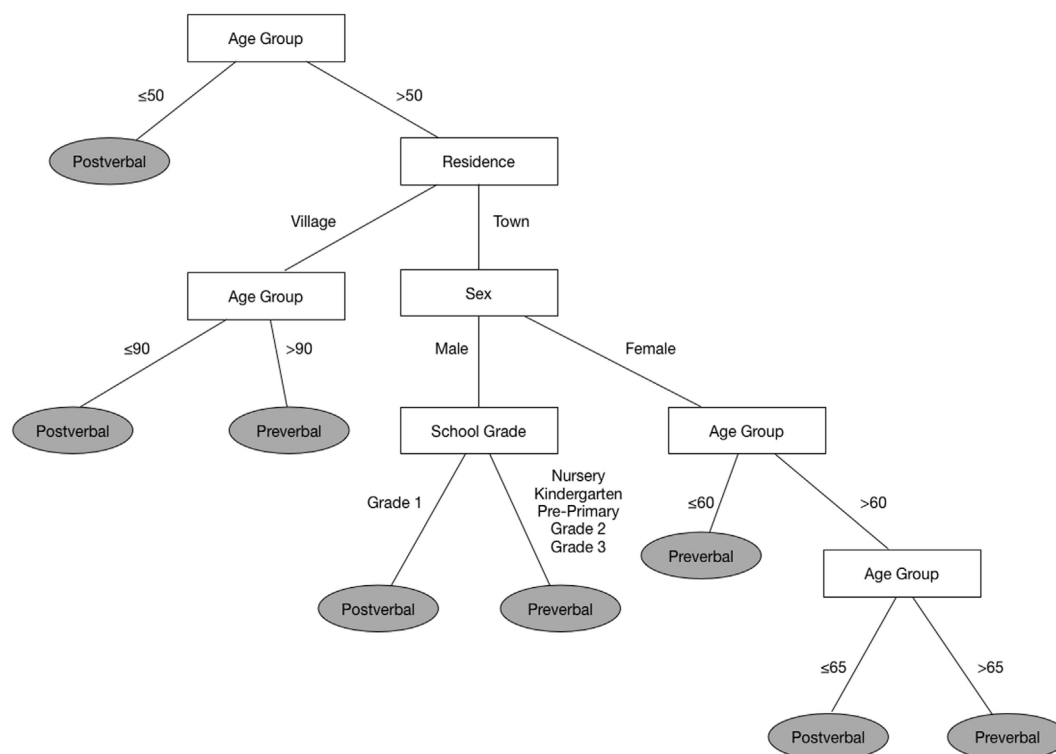


FIGURE 5 | C5.0 decision tree.

(H3) assumes an effect due to the interaction of social factors. It has become evident by now that social factors interact. But let us examine another case of this interaction. Strikingly, in 90- to 100-month-old children (around 8 years), preverbal clitic placement increases dramatically, without obvious linguistic reason. Observe that the influence of sociolinguistic factors becomes more evident after 65 months, at around age 5 years 5 months. Boys employ enclisis to a greater degree than girls. Similarly, speakers who live in cities employ more standard and fewer marked forms than speakers who live in rural places. This finding suggests that girls adhere to standard forms from very early on, and they are more sensitive to norms and to prestigious and recognized forms [for further discussion, see, e.g., Trudgill (1972) or Labov (2001)].

Most importantly, at around 50 months, the *place of residence* plays a significant role (these effects are represented as a tree node for place of residence, which dominates gender). However, after 75 months of age, children begin to comprehend and appreciate gender roles, and to adhere to these; thus, they make language choices, which are represented as particular choices of clitic placement and acknowledge the social and linguistic environment. At around age of 8 years 4 months (100 months), children can discriminate standard from non-standard forms and appreciate the functions of these. In addition, they also have operational knowledge, namely to apply these distinctions in their everyday sociolinguistic practices and interactions [see also the studies by Reid (1978), Romaine (1978), or Payne (1980), among others, on early work regarding the effects of place of residence on children's linguistic choices].

When parallel structures exist in children's sociolinguistic environment (such as SMG, CSG, as well as basilectal and acrolectal varieties of CG), learning the target grammar involves a process of selection. Children develop language-specific practices as they grow up. At first, children follow the language of their caregivers, that is, the linguistic input found in children's immediate linguistic environment. But as they grow up and develop social skills, they become more communicative and (language-)competent members of society. In other words, they become more influenced by their immediate community: at first the wider family and, increasingly, friends from peer groups. This environment influences their linguistic behavior as they are reflected in the specific choices children make in clitic placement. Studies that include adolescents and preadolescents show a very similar pattern. Preadolescents tend to adopt more innovative variants, which peak in adolescence (Labov, 1994; Tagliamonte, 2012). This tendency affects most aspects of grammar, from phonology (Ash, 1982; Tagliamonte, 2012) to morphosyntax and discourse (Trudgill, 1974; Ash, 1982).

Overall, variation is inherent in monolingual and multilingual environments. Individual speakers acquire the characteristic frequency for particular variables from their caregivers. The behavior of speakers and that of the language community do not remain stable throughout speakers' lifetime (Tagliamonte, 2012, 49f). The frequency of different variables depends on the different stages of a speaker's life; it may increase in adolescence and even undergo reorganization (Labov, 1994). However, by late adolescence, a speaker's linguistic system stabilizes, and from that point onward, it is maintained for the rest of his or her life (Kirkham and

Moore, 2013). Speakers between 30 and 55 years tend to employ more standard and fewer marked forms (Labov, 1994). In older age, non-prestigious forms may reappear as speakers relax and detach themselves from the need to confront to society's demands (Labov, 1994; Tagliamonte, 2012). Future research on bilectal linguistic behavior in Cyprus through the lifetime should address will undoubtedly shed further light on these issues.

## CONCLUSION

The study examined the acquisition of clitic placement (preverbal vs. postverbal) in Cypriot Greek, which is characterized by diglossia/bilectalism. We raised the following questions: How does the acquisition of clitics develop over time? How do sociolinguistic factors such as gender (male/female) and habitation setting (urban vs. rural) determine clitic placement? How does schooling affect clitic placement? The results of the study presented social factors that interact during language acquisition, especially postcritical period when the two emerging grammars seem to compete.

By employing a learning and clustering approach, the analysis provides a perhaps better understanding of these interactions, captured by the Socio-Syntax of Development Hypothesis,

which can be understood as the sociolinguistic trigger for the observed grammatical competition. Further research currently carried out under the first author's supervision investigates other aspects of bilectal grammar aiming to tie these to closer to both executive control abilities [cf. Antoniou et al. (2016), but also Garraffa et al. (2015)] and, more generally, a gradient scale of multilingualism (cf. Grohmann, 2014b; Grohmann and Kambanaros, 2015).

## ETHICS STATEMENT

According to the local legislation in Cyprus, ethical approval was not required for this type of study. The study was, however, conducted in accordance with the declaration of Helsinki, and written informed consent was obtained from the parents of each participant.

## AUTHOR CONTRIBUTIONS

KG envisioned the project, managed the research, and revised the paper. EP was responsible for data collection, compilation, and entry. CT ran the statistical analysis and wrote a first draft of the paper. Subsequently all authors worked on refining and revising the text. All authors approved the final version.

## REFERENCES

- Agathocleous, M. (2012). *The role of the Social Environment in the Production of Clitics in Cypriot Greek within the COST Action A33 Testing Tool*. MA thesis, University of Cyprus, Nicosia.
- Agathocleous, M., Charalambous, A., Papadopoulou, E., and Grohmann, K. K. (2014). "The role of the social environment on linguistic development," in *Developments in the Acquisition of Clitics*, eds K. K. Grohmann and T. Neokleous (Newcastle upon Tyne: Cambridge Scholars Publishing), 276–291.
- Du Bois, J. A. (1985). "Competing motivations," in *Iconicity in Syntax*, ed. J. Haiman (Amsterdam: John Benjamins), 343–366.
- Duarte, I., and Matos, G. (2000). "Romance clitics and the minimalist program," in *Portuguese Syntax*, ed. J. Costa (New York: Oxford University Press), 116–142.
- Eckert, P. (1999). *Linguistic Variation as Social Practice*. Oxford: Blackwell.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., and Pethick, S. J. (1994). Variability in early communicative development. *Soc. Res. Child Dev.* 59, 1–173.
- Ferguson, C. (1959). Diglossia. *Word* 15, 325–340. doi:10.1080/00437956.1959.1659702
- Garraffa, M., Beveridge, M., and Sorace, M. (2015). Linguistic and cognitive skills in Sardinian-Italian bilingual children. *Front. Psychol.* 6:1898. doi:10.3389/fpsyg.2015.01898
- Goutsos, D., and Karyolemou, M. (2004). Introduction. *Int. J. Soc. Lang.* 168, 1–17. doi:10.1515/ijsl.2004.031
- Grohmann, K. K. (2011). "Some directions for the systematic investigation of the acquisition of Cypriot Greek: A new perspective on production abilities from object clitic placement," in *The Development of Grammar: Language Acquisition and Diachronic Change*, eds E. Rinke and T. Kupisch (Amsterdam: John Benjamins), 179–203.
- Grohmann, K. K. (2014a). "CAT research on object clitic placement: Where we are now?" in *Developments in the Acquisition of Clitics*, eds K. K. Grohmann and T. Neokleous (Newcastle upon Tyne: Cambridge Scholars Publishing), 1–40.
- Grohmann, K. K. (2014b). Towards comparative bilingualism. *Linguist. Approaches Biling.* 4, 336–341. doi:10.1075/lab.4.3.06gro
- Grohmann, K. K., and Kambanaros, M. (2015). The gradience of multilingualism in typical and impaired language development: positioning bilectalism within comparative bilingualism. *Front. Psychol.* 7:37. doi:10.3389/fpsyg.2016.00037
- Grohmann, K. K., and Leivada, E. (2012). "Interface ingredients of dialect design bi-x, socio-syntax of development, and the grammar of Cypriot Greek," in *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces*, ed. A. M. Di Sciullo (Amsterdam: John Benjamins), 239–262.
- Cohen, W. W. (1995). "Fast effective rule induction," in *Proceedings of the Twelfth International Conference on Machine Learning*, eds A. Prieditis and S. J. Russell (San Francisco, CA: Kaufmann), 115–123.
- Costa, J., and Lobo, M. (2014). "Crosslinguistic variation in the acquisition of null categories: Portuguese vs. Japanese," in *Developments in the Acquisition of Clitics*, eds K. K. Grohmann and T. Neokleous (Newcastle upon Tyne: Cambridge Scholars Publishing), 276–291.
- Du Bois, J. A. (1985). "Competing motivations," in *Iconicity in Syntax*, ed. J. Haiman (Amsterdam: John Benjamins), 343–366.
- Duarte, I., and Matos, G. (2000). "Romance clitics and the minimalist program," in *Portuguese Syntax*, ed. J. Costa (New York: Oxford University Press), 116–142.
- Eckert, P. (1999). *Linguistic Variation as Social Practice*. Oxford: Blackwell.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., and Pethick, S. J. (1994). Variability in early communicative development. *Soc. Res. Child Dev.* 59, 1–173.
- Ferguson, C. (1959). Diglossia. *Word* 15, 325–340. doi:10.1080/00437956.1959.1659702
- Garraffa, M., Beveridge, M., and Sorace, M. (2015). Linguistic and cognitive skills in Sardinian-Italian bilingual children. *Front. Psychol.* 6:1898. doi:10.3389/fpsyg.2015.01898
- Goutsos, D., and Karyolemou, M. (2004). Introduction. *Int. J. Soc. Lang.* 168, 1–17. doi:10.1515/ijsl.2004.031
- Grohmann, K. K. (2011). "Some directions for the systematic investigation of the acquisition of Cypriot Greek: A new perspective on production abilities from object clitic placement," in *The Development of Grammar: Language Acquisition and Diachronic Change*, eds E. Rinke and T. Kupisch (Amsterdam: John Benjamins), 179–203.
- Grohmann, K. K. (2014a). "CAT research on object clitic placement: Where we are now?" in *Developments in the Acquisition of Clitics*, eds K. K. Grohmann and T. Neokleous (Newcastle upon Tyne: Cambridge Scholars Publishing), 1–40.
- Grohmann, K. K. (2014b). Towards comparative bilingualism. *Linguist. Approaches Biling.* 4, 336–341. doi:10.1075/lab.4.3.06gro
- Grohmann, K. K., and Kambanaros, M. (2015). The gradience of multilingualism in typical and impaired language development: positioning bilectalism within comparative bilingualism. *Front. Psychol.* 7:37. doi:10.3389/fpsyg.2016.00037
- Grohmann, K. K., and Leivada, E. (2012). "Interface ingredients of dialect design bi-x, socio-syntax of development, and the grammar of Cypriot Greek," in *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces*, ed. A. M. Di Sciullo (Amsterdam: John Benjamins), 239–262.
- Agathocleous, M. (2012). *The role of the Social Environment in the Production of Clitics in Cypriot Greek within the COST Action A33 Testing Tool*. MA thesis, University of Cyprus, Nicosia.
- Agathocleous, M., Charalambous, A., Papadopoulou, E., and Grohmann, K. K. (2014). "The role of the social environment on linguistic development," in *Developments in the Acquisition of Clitics*, eds K. K. Grohmann and T. Neokleous (Newcastle upon Tyne: Cambridge Scholars Publishing), 276–291.
- Agouraki, Y. (1997). "On the enclisis/proclisis alternation," in *Proceedings of the 2nd International Conference on Greek Linguistics*, eds G. Drachman, A. Malikouti Drachman, J. Fykias, and C. Klidi (Graz: W. Neugebauer Verlag), 393–404.
- Agouraki, Y. (2001). "The position of clitics in Cypriot Greek," in *Proceedings of the First International Conference of Modern Greek Dialects and Linguistic Theory (Patras, Greece, Oct. 12–14, 2000)*, eds A. Ralli, B. D. Joseph, and M. Janse (Patras: University of Patras), 1–17.
- Andreou, M. (2015). *The Effects of Bilingualism on Verbal and Non-Verbal Cognition: The Micro- and Macro-Structure of Narratives in the Weak and the Dominant Language of a Bilingual Child [Dissertation]*. Thessaloniki: Aristotle University of Thessaloniki.
- Antoniou, K., Grohmann, K. K., Kambanaros, M., and Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition* 149, 18–30. doi:10.1016/j.cognition.2015.12.002
- Arvaniti, A. (2010). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *Mediterr. Lang. Rev.* 17, 15–45.
- Ash, S. (1982). *The Vocalization of /l/ in Philadelphia [Dissertation]*. Philadelphia, PA: University of Pennsylvania.
- Barbu, S., Nardy, A., Chevrot, J. P., Guellai, B., Glas, L., Juhel, J., et al. (2015). Sex differences in language across early childhood: family socioeconomic status does not impact boys and girls equally. *Front. Psychol.* 6:1874. doi:10.3389/fpsyg.2015.01874
- Cazden, C. B. (1976). "Play with language and metalinguistic awareness: One dimension of language experience," in *Play: Its Role in Development and Evolution*, eds J. S. Bruner, A. Jolly, and K. Sylva (New York: Basic Books), 603–608.
- Chatzikyriakidis, S. (2010). *Clitics in Four Dialects of Modern Greek: A Dynamic Account [Dissertation]*. London: King's College.
- Chatzikyriakidis, S. (2012). A dynamic account of clitic positioning in Cypriot Greek. *Lingua* 122, 642–672. doi:10.1016/j.lingua.2012.01.008

- Grohmann, K. K., Theodorou, E., Pavlou, N., Leivada, E., Papadopoulou, E., Martinez-Ferreiro, S., et al. (2012). "The development of clitic placement in Cypriot Greek and the romance connection," in *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages*, eds P. Prévost, L. Tuller, and R. Zebib (Newcastle upon Tyne: Cambridge Scholars Publishing), 128–152.
- Habib, R. (2016). Bidirectional linguistic change in rural child and adolescent language in Syria. *Dialectologia* 16, 117–141.
- Hadjioannou, X., Tsiplakou, S., and Kappler, M. (2011). Language policy and language planning in Cyprus. *Curr. Issues Lang. Plan.* 12, 503–569. doi:10.1080/14664208.2011.629113
- Heycock, C., Sorace, A., Hansen, Z., and Wilson, F. D. (2013). Acquisition in variation (and vice versa): V-to-T in Faroese children. *Lang. Acquis.* 20, 5–22. doi:10.1080/10489223.2012.738741
- Hyde, J. S. (2005). The gender similarities hypothesis. *Am. Psychol.* 60, 581–592. doi:10.1037/0003-066X.60.6.581
- Kirkham, S., and Moore, E. (2013). "Adolescence," in *The Handbook of Language Variation and Change*, eds J. Chambers and N. Schilling (Oxford: Wiley-Blackwell), 277–296.
- Kroch, A. (1994). "Morphosyntactic variation," in *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, eds K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, and E. Zeinfeld (Chicago: Chicago Linguistics Society), 180–201.
- Kroch, A., and Taylor, A. (2000). "Verb-object order in Middle English," in *Diachronic Syntax: Models and Mechanisms*, eds S. Pintzuk, G. Tsoulas, and A. Warner (Oxford: Oxford University Press), 132–163.
- Labov, W. (1994). *Principles of Linguistic Change, Vol. 1: Internal Factors*. Oxford: Blackwell.
- Labov, W. (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford: Blackwell.
- Lascaratou, C. (1998). "Basic characteristics of Modern Greek word order," in *Constituent Order in the Languages of Europe*, ed. A. Siewierska (Berlin: Mouton de Gruyter), 151–171.
- Legate, J., and Yang, C. (2007). Morphosyntactic learning and the development of tense. *Lang. Acquis.* 14, 315–344. doi:10.1080/10489220701471081
- Leivada, E., and Grohmann, K. K. (in press). "Language acquisition in bilectal environments: competing motivations, metalinguistic awareness, and the socio-syntax of development hypothesis," in *Variation in Language Acquisition*, eds M. Katerbow and G. de Vogelaer (Amsterdam: John Benjamins).
- Leivada, E., Kambanaros, M., Taxitari, L., and Grohmann, K. K. (in press). (Meta) linguistic abilities of bilectal educators: the case of Cyprus. *Int. J. Biling. Educ. Biling.*
- Leivada, E., Mavroudi, P., and Epistithiou, A. (2010). "Metalanguage or bidialectism? Acquisition of clitic placement by Hellenic Greeks, Greek Cypriots and binationals in the diglossic context of Cyprus," in *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics 2010*, ed. A. Botinis (Athens: ISCA and the University of Athens), 97–100.
- Leivada, E., Papadopoulou, E., Kambanaros, M., and Grohmann, K. K. (2017). The influence of bilectalism and non-standardization on the perception of native grammatical variants. *Front. Psychol.* 8:205. doi:10.3389/fpsyg.2017.00205
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley & Sons.
- Lightfoot, D. (1999). *The Development of Language: Acquisition, Change, and Evolution*. Oxford: Blackwell.
- Lorenzo, G. (1994). Optionality in the placement of Asturian clitics. *Catalan Work. Pap. Linguist.* 4.1, 93–120.
- Maccoby, E. E., and Jacklin, C. N. (1974). *The Psychology of Sex Differences*. Stanford, CA: Stanford University Press.
- Marinis, T. (2000). The acquisition of clitic objects in Modern Greek: single clitics, double clitics, clitic left dislocation. *ZAS Pap. Linguist.* 15, 260–283.
- Nesdale, D., and Rooney, R. (1996). Evaluations and stereotyping of accented speakers by pre-adolescent children. *J. Lang. Soc. Psychol.* 15, 133–154. doi:10.1177/0261927X960152002
- Newton, B. E. (1972). *Cypriot Greek: Its Phonology and Inflections*. The Hague: Mouton.
- Pappas, P. (2012). "An empirical perspective on Cypriot clitics," in *Neolliniki dialektologia [Modern Greek Dialectology]*, Vol. 6, eds C. Bezandakou and K. Minas (Athens: KSILK), 391–413.
- Pappas, P. (2014). Exceptional clitic placement in Cypriot Greek: results from an MET study. *J. Greek Linguist.* 14, 190–211. doi:10.1163/15699846-01402002
- Payne, A. C. (1980). "Factors controlling the acquisition of the Philadelphia dialect by out-of-state children," in *Locating Language in Time and Space*, ed. W. Labov (New York: Academic Press), 143–178.
- Petinou, K., and Terzi, A. (2002). Clitic misplacement among normally developing and SLI children and the status of Infl heads. *Lang. Acquis.* 10, 1–28. doi:10.1207/S15327817LA1001\_1
- Philippaki-Warbuton, I. (1985). Word order in Modern Greek. *Trans. Phil. Soc.* 83, 113–143.
- Raposo, E. (1986). "On the null object in European Portuguese," in *Studies in Romance Linguistics*, eds O. Jaeggli and C. Silva Corvalán (Dordrecht: Foris), 373–390.
- R Core Team. (2012). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org>
- Reid, E. (1978). "Social and stylistic variation in the speech of children: some evidence from Edinburgh," in *Sociolinguistic Patterns in British English*, ed. P. Trudgill (London: Arnold), 158–172.
- Revithiadou, A. (2006). Prosodic filters on syntax: an interface account of second position clitics. *Lingua* 116, 79–111. doi:10.1016/j.lingua.2004.08.017
- Revithiadou, A., and Spyropoulos, V. (2008). Greek object clitic pronouns: a typological survey of their grammatical properties. *STUF* 61, 39–53.
- Romaine, S. (1978). "Post-vocalic/r/ in Scottish English: sound change in progress," in *Sociolinguistic Patterns in British English*, ed. P. Trudgill (London: Arnold), 144–157.
- Roussou, A., and Tsimpli, I. M. (2006). On Greek VSO again. *J. Linguist.* 42, 317–354. doi:10.1017/S002226706003914
- Rowe, C., and Grohmann, K. K. (2013). Discrete bilectalism: towards co-overt prestige and diglossic shift in Cyprus. *Int. J. Soc. Lang.* 224, 119–142.
- Rowe, C., and Grohmann, K. K. (2014). Canaries in a coal mine: native speakerhood and other factors as predictors of moribundity, death, and diglossic shift in Cypriot Greek. *Mediterr. Lang. Rev.* 21, 121–142.
- Russell, S. J., and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, 2nd Edn. Englewood Cliffs, NJ: Prentice Hall.
- Siegel, J. (2010). *Second Dialect Acquisition*. Cambridge: Cambridge University Press.
- Sophocleous, A. (2011). Two languages in the classroom: the inconsistency between national and local objectives in formal education in Cyprus. *J. Lang. Identity Educ.* 10, 266–281. doi:10.1080/15348458.2011.598129
- Statistical Service of Cyprus (CYSTAT). (2011). *Population Census. Population and Social Conditions*. Nicosia. Available at: [http://www.cystat.gov.cy/mof/cystat/statistics.nsf/populationcondition\\_22main\\_en?OpenForm&sub=2&sel=2#](http://www.cystat.gov.cy/mof/cystat/statistics.nsf/populationcondition_22main_en?OpenForm&sub=2&sel=2#)
- Tagliamonte, S. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden, MA: Wiley-Blackwell.
- Terzi, A. (1999a). Clitic combinations, their hosts and their ordering. *Nat. Lang. Linguist. Theory* 17, 85–121. doi:10.1023/A:1006075626603
- Terzi, A. (1999b). "Cypriot Greek clitics and their positioning restrictions," in *Studies in Greek Syntax*, eds A. Alexiadou, G. Horrocks, and M. Stavrou (Dordrecht: Kluwer), 227–240.
- Theodorou, E. (2007). *Phonetic Development of CG-Speaking Toddlers Ages 24 to 36 Months: A Longitudinal Study*. M.Sc. thesis, University of Sheffield, Sheffield.
- Theodorou, E. (2013). *Specific Language Impairment in Cypriot Greek: Diagnostic and Experimental Investigations*. PhD dissertation, University of Cyprus, Nicosia.
- Theodorou, E., and Grohmann, K. K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua* 161, 144–158. doi:10.1016/j.lingua.2014.11.011
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Lang. Soc.* 1, 179–195. doi:10.1017/S004740450-0000488
- Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Tsimpli, I. M. (2014). Early, late or very late: timing acquisition and bilingualism. *Linguist. Approaches Biling.* 4, 393–402. doi:10.1075/lab.4.3.01tsi
- Tsiplakou, S. (2006). The emperor's old clothes: linguistic diversity and the redefinition of literacy. *Int. J. Humanit.* 2, 2345–2352.
- Tsiplakou, S. (2009). "Language alternation, competing grammars and inter-language: another facet of the Cyprus problem," in *Proceedings of the 8th*

- International Conference on Greek Linguistics*, eds G. Giannakis, M. Baltazani, G. Xydopoulos, and T. Tsangalidis (Ioannina: University of Ioannina), 1195–1209.
- Tsiplakou, S. (2014). How mixed is a 'mixed' system? The case of the Cypriot Greek *koiné*. *Linguist. Var.* 14, 161–178. doi:10.1075/lv.14.1.07tsi
- Vangsnes, ØA., Söderlund, G. B. W., and Blekesaune, M. (2017). The effect of bidialectal literacy on school achievement. *Int. J. Biling. Educ. Biling.* doi:10.1080/13670050.2015.1051507
- Varlokosta, S., Belletti, A., Costa, J., Friedmann, N., Gavarró, A., Grohmann, K. K., et al. (2016). A cross-linguistic study of the acquisition of clitic and pronoun production. *Lang. Acquis.* 23, 1–26. doi:10.1080/10489223.2015.1028628
- Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: a critical review. *Brain Lang.* 108, 175–183. doi:10.1016/j.bandl.2008.07.001
- Wolfram, W. (1994). "Bidialectal literacy in the United States," in *Adult Bilinguality in the United States*, ed. D. Spener (Washington, DC; McHenry, IL: Center for Applied Linguistics and Delta Systems), 71–88.
- Yang, C. (2000). Internal and external forces in language change. *Lang. Var. Change* 12, 231–250. doi:10.1017/S0954394500123014

**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 © 2017 Grohmann, Papadopoulou and Themistocleous. 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.



# The Influence of Bilectalism and Non-standardization on the Perception of Native Grammatical Variants

Evelina Leivada<sup>1,2\*</sup>, Elena Papadopoulou<sup>2,3</sup>, Maria Kambanaros<sup>1,2</sup> and Kleanthes K. Grohmann<sup>2,4</sup>

<sup>1</sup> Department of Rehabilitation Sciences, Cyprus University of Technology, Limassol, Cyprus, <sup>2</sup> Cyprus Acquisition Team, University of Cyprus, Nicosia, Cyprus, <sup>3</sup> Department of Humanities, European University Cyprus, Nicosia, Cyprus,

<sup>4</sup> Department of English Studies, University of Cyprus, Nicosia, Cyprus

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University, UK

### Reviewed by:

Urs Maurer,  
The Chinese University of Hong Kong,  
Hong Kong  
Aritz Iruztun,  
Centre National de la Recherche  
Scientifique (CNRS), France

### \*Correspondence:

Evelina Leivada  
evelina@biolinguistics.eu

### Specialty section:

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

**Received:** 02 October 2016

**Accepted:** 01 February 2017

**Published:** 20 February 2017

### Citation:

Leivada E, Papadopoulou E,  
Kambanaros M and Grohmann KK  
(2017) The Influence of Bilectalism  
and Non-standardization on  
the Perception of Native Grammatical  
Variants. *Front. Psychol.* 8:205.  
doi: 10.3389/fpsyg.2017.00205

Research in speakers of closely related varieties has shown that bilectalism and non-standardization affect speakers' perception of the variants that exist in their native languages in a way that is absent from the performance of their monolingual peers. One possible explanation for this difference is that non-standardization blurs the boundaries of grammatical variants and increases grammatical fluidity. Affected by such factors, bilectals become less accurate in identifying the variety to which a grammatical variant pertains. Another explanation is that their differential performance derives from the fact that they are competent in two varieties. Under this scenario, the difference is due to the existence of two linguistic systems in the course of development, and not to how close or standardized these systems are. This study employs a novel variety-judgment task in order to elucidate which of the two explanations holds. Having administered the task to monolinguals, bilectals, and bilinguals, including heritage language learners and L1 attriters, we obtained a dataset of 16,245 sentences. The analysis shows differential performance between bilectal and bilingual speakers, granting support for the first explanation. We discuss the role of factors such as non-standardization and linguistic proximity in language development and flesh out the implications of the results in relation to different developmental trajectories.

**Keywords:** bilectalism, dialect, grammatical variants, non-standardization

## INTRODUCTION

Linguistic research has shown that non-standard varieties allow for greater grammatical fluidity in a way that blurs the boundaries across them and affects speakers' perception of whether a specific variant belongs to their linguistic repertoire or not (Cheshire and Stein, 1997; Henry, 2005). Non-standardization affects not only cross-linguistic boundaries but also the norms of acceptability that define variants (Milroy, 2001). This, in turn, affects speakers' perception and ultimate performance of grammatical variants in their native variety or varieties (Henry, 2005; Papadopoulou et al., 2014). A second important factor that affects linguistic development is variation in the input, as happens when the linguistic environment involves exposure to more than one language. Bilingual speakers benefit from the cognitive advantages of bilingualism, which have an impact on the processing mechanisms that are active during the acquisition process. For example, bilingualism strengthens

the development of the attentional control abilities with the effect persisting throughout the lifetime (Bialystok et al., 2004; Laka, 2012). Linguistic proximity across the different languages a child is exposed to is a third key factor that affects bi- or multilingual development and cross-linguistic transfer (Grohmann, 2014; Garraffa et al., 2015; Westergaard et al., 2016).

The interaction of these three factors ultimately invests the linguistic development of speakers that become exposed to two *closely related varieties*—henceforth, ‘(discrete) bilectal’ speakers (Rowe and Grohmann, 2013)—with a cluster of unique properties that can be described as inherent to the notion of ‘dialect design’ (Grohmann and Leivada, 2012): (i) blurred boundaries of grammatical variants, (ii) dialect continua and the emergence of intermediate speech repertoires (Cornips, 2006), possible lack of codification, and a prescriptive notion of correctness that interferes with speakers’ perception of their own linguistic repertoire (Henry, 2005; Grohmann and Leivada, 2012).

A striking result from the field of experimental linguistics relates to the finding that a native speaker may judge a certain variant or form to be completely unacceptable, but be recorded producing it in their own speech (Cornips and Poletto, 2005). If this is true in cases of mono- or bilingual development, in bilectal development it involves non-official/-codified—and as such, more fluid—varieties, rendering an even greater degree of discrepancy between speakers’ introspective judgments about their repertoire and the actual linguistic repertoire itself. Even in the absence of closely related varieties and the dialect design, bilingualism may leave its imprint on speakers’ performance in acceptability judgment tasks.<sup>1</sup> This results in observing differential performance across mono- and bilingual populations in both off-line measures (e.g., the higher acceptance rate of over-regularizations in bilinguals reported in Jacobson and Cairns, 2008) and on-line measures (e.g., the slower reaction times of balanced bilinguals in Foursha et al., 2006). Experiments measuring acceptability judgment using event-related potentials have also shown lower levels of performance in bilinguals compared to monolinguals for some tasks, and a different distribution of activation across the two groups (Moreno et al., 2010).

**Table 1** presents methods and outcomes of acceptability judgment experiments in adult, bilingual populations.<sup>2</sup> Evidently, bilinguals perform similarly to monolinguals only in some tasks and may differ in on-line responses. Bilingualism and variation in the input can affect speakers’ performance in evaluating the

structures that (do not) form part of their repertoire. Observing such differences between monolingual and bilingual speakers, the question that arises in this context is whether the greater degree of fluidity that bilectalism, non-standardization, and linguistic proximity entail leads to a performance that is different from that of monolingual and/or bilingual speakers. To this date, no study has compared the performance of monolinguals, bilinguals, and bilectals in a task that measures the ability to identify variants that belong (or not) to their native variety or varieties.

A recent study investigating the linguistic profile of bilectal speakers of Cypriot Greek (CG) and Standard Modern Greek (SMG) in a written variety-judgment task that superimposed dialectal elements from CG on SMG stimuli revealed important differences between the two groups of speakers across all levels of linguistic analysis (Leivada et al., forthcoming). Despite the fact that this study provided a novel comparison between monolingual and bilectal speakers of two different varieties of Greek, it cannot answer the question of differential performance across monolinguals, bilinguals, and bilectals for two reasons. First, the study tested only school teachers, not the general population. The reasoning behind this was sociolinguistic in nature.

Specifically, Cyprus involves a state of diglossia, with SMG being the sociolinguistically ‘H(igh)’- and CG the ‘L(ow)’-variety. SMG, the official language of the state, is the language used in education and other formal settings.<sup>3</sup> Pavlou and Papapavlou’s (2004) investigation of dialect use in education puts emphasis on the fact that the language of instruction in Cyprus is SMG and the majority of the textbooks are produced in Greece. As a result, the official policy “forces teachers to adopt as part of their teaching methodology the following principles: SMG should be the exclusive code of instruction and of general use in class; and students should be “corrected” when *using dialect words*, when pronouncing words with a *Cypriot accent*, and when using phonological rules that are part of the *phonological system of GC*” (p. 250; emphasis added).

Despite what the official policy requires, numerous studies have revealed interference of CG in oral and/or written discourse from the students’ perspectives (Pavlou and Christodoulou, 2001; Ioannidou, 2002; Papapavlou and Yiakoumetti, 2003). Experiments focusing on teachers’ output and possible CG interference in *their* repertoire are scarce (Karyolemou, 2006). This is the topic of Leivada et al. (forthcoming). The underlying assumption is that teachers are in a better position than anyone else in Cyprus to demonstrate advanced linguistic performance in the standard language—SMG. In fact, their command of the language should be comparable to that of Hellenic Greeks, monolingual native speakers of SMG, given that they are educated and professionally trained to teach in this variety. The results of Leivada et al. (forthcoming) revealed differences in the performance of bilectal, Greek Cypriot teachers compared to monolingual, Hellenic Greek teachers, as the former were

<sup>1</sup>Throughout this work, the term ‘bilingualism’ denotes competence in two different languages, whereas the term ‘bilectalism’ is employed when there is great structural proximity between the two varieties of a speaker. We also prefer use of the term ‘bilectal’ over ‘bidialectal’ for purposes of precision: The H-variety in many diglossic speech communities is a standard language, which is a superposed variety and not a dialect.

<sup>2</sup>We use the term ‘acceptability’ rather than ‘grammaticality,’ although many studies employ both, sometimes even interchangeably or as if there was a measurable contrast (see **Table 1**). There simply is no list of linguistic stimuli that are grammatical *in and of themselves*, hence we talk about *acceptability*. For further discussion, see also Bard et al. (1996), Schütze (1996), Keller (2000), or Sprouse et al. (2013) and the broader issues laid out in Sprouse and Almeida (2013).

<sup>3</sup>For the linguistic reality of Cyprus, see Rowe and Grohmann (2013, 2014) and for attitudes toward CG in the classroom environment see Sophocleous and Wilks (2010), including relevant references cited there.

**TABLE 1 | Relevant studies with adult populations.**

| Study                      | Languages                                  | Method                                                                                                                                                                                                                                                                                                                                                                                                     | Outcome                                                                                                                                                                                                                                                                                                     |
|----------------------------|--------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Foursha et al. (2006)      | English–Spanish                            | Participants were asked to press one key if they thought the sentence they heard was grammatical and another key, if they thought the sentence was ungrammatical.                                                                                                                                                                                                                                          | (1) Monolinguals outperformed bilinguals on converging over conflicting sentences for very few sentence pairs.<br>(2) Bilinguals produced slower overall reaction times.<br>(3) Bilinguals and monolinguals did not present different patterns of performance for the conflicting and converging sentences. |
| Jacobson and Cairns (2008) | English–Spanish                            | Participants were asked to report acceptability on a binary scale after being orally presented with a sentence.                                                                                                                                                                                                                                                                                            | (1) Monolinguals overwhelmingly rejected hearing or using over-regularizations of the sort <i>caught</i> vs. <i>catched</i> .<br>(2) Bilinguals reported hearing and using some of the over-regularizations to a greater degree than monolinguals.                                                          |
| Moreno et al. (2010)       | English–French/Hebrew/<br>Romanian/Russian | Participants were presented with the sentences word-by-word and were asked to press one of the two buttons, answering “yes” if the sentence was good and “no” if there was something wrong with the sentence. The <i>acceptability</i> task asked participants to identify errors in grammar or meaning. The <i>grammaticality</i> task participants indicated only errors in grammar. ERPs were recorded. | (1) In the acceptability task, bilinguals were less accurate than monolinguals.<br>(2) In the grammaticality task, the two groups showed a comparable level of accuracy.<br>(3) Bilinguals generated smaller P600 amplitude and a more bilateral distribution of activation than monolinguals.              |

significantly less accurate than the latter in identifying dialectal elements and correctly classifying the test stimuli as SMG or not.

The second reason that Leivada et al. (forthcoming) cannot address the issue of differential performance across monolinguals, bilinguals, and bilectals is that no bilingual group was tested in that study, hence no comparison of bilinguals with bilectals is possible. The present experiment aims to fill this gap in the literature, through investigating potential differences in the performance of monolingual, bilingual, and bilectal populations in a written task and identifying the factors that affect this performance.

## Aims and Predictions

The present study aims to answer the question of whether the differential performance across monolinguals and bilectals reported in Leivada et al. (forthcoming) is the result of being exposed to more than one linguistic system in the course of development. If so, the prediction is that bilectals will perform like bilinguals, since both have exposure to more than one linguistic system. However, if factors inherent to the dialect design such as non-standardization and linguistic proximity between the two varieties come into play, one expects bilectals to perform differently than bilinguals and monolinguals. In sum, the starting point for investigating speakers’ perception of native grammatical variants, can be summarized in the following two possible causes of differential performance across groups.

- (1) Non-standardization of CG and its close linguistic proximity to SMG blur the boundaries of grammatical variants and increase grammatical fluidity. Consequently, Greek Cypriots’ perception of their linguistic repertoire is affected by such factors. They become less accurate in

spotting the dialectal elements they are presented with in variety judgment task and classifying the test stimuli as CG or SMG.

- (2) The differential performance is not due to the existence of two closely related varieties but to the fact that Greek Cypriots are competent in the two varieties. Put differently, their performance is related to the *existence* of two varieties and not to which (and how close or standardized) these two varieties are. As mentioned already, the relevant literature reports that bilinguals perform differently than monolinguals in some acceptability judgment tasks, and this difference could plausibly extend to bilectals.

In sum, bilectals have exposure to both varieties so, in principle, one could think they are in a more privileged position to correctly identify the test stimuli compared to the other groups. We do not favor this possibility, but we acknowledge it as valid hypothesis among others. The aim is to empirically (dis)confirm this idea, especially since it *could* be the case that exposure to more than one language leaves its imprint in the same way across bilingual and bilectal populations. To the best of our knowledge, no other study has compared monolinguals, bilectals, and different type of bilinguals, hence no other study has already ruled out this possibility. If this privilege is not reflected in the performance of bilectals, this could mean that other factors intervene and cloud their ability to identify the grammatical variants that are part of their linguistic repertoires. In this case, the performance of the bilingual group will be the control that can elucidate whether these other factors boil down to (1) or (2).

## Participants

A total of 361 participants took part in this study, divided into four groups:

**TABLE 2 | All groups.**

| Groups                     |                         | N                | Age range (in years) | Mean age (in years) (SD) |
|----------------------------|-------------------------|------------------|----------------------|--------------------------|
| Group I <sub>HG</sub>      |                         | 100 (77 females) | 19–67                | 32.12 (10.2)             |
| Group II <sub>GC</sub>     |                         | 100 (49 females) | 18–72                | 32.16 (12.9)             |
| Group III <sub>GC-GR</sub> |                         | 61 (35 females)  | 19–65                | 33.82 (11.57)            |
| Group IV <sub>BI</sub>     | Simultaneous bilinguals | 33 (23 females)  | 20–59                | 36.6 (10.09)             |
|                            | Heritage speakers       | 25 (19 females)  | 22–56                | 38.56 (8.51)             |
|                            | L1 attriters            | 42 (30 females)  | 26–80                | 44.45 (12.03)            |
|                            | Overall                 | 100 (71 females) | 20–80                | 40.39 (11.09)            |

SD stands for Standard Deviation.

- (i) Hundred Hellenic Greek monolingual speakers of SMG who have acquired language in the monolingual environment of Greece (Group I<sub>HG</sub>; see Table 2 for further breakdown);
- (ii) Hundred Greek Cypriot bilectals who have been exposed to both SMG and CG in the bilectal context of Cyprus (Group II<sub>GC</sub>);
- (iii) Sixty-one Greek Cypriot bilectals who have acquired SMG and CG in the bilectal environment of Cyprus and have lived in Greece for more than 1 year during adulthood, most of them studying toward a 4-year university degree (Group III<sub>GC-GR</sub>, mean age in Greece: 3.6 years); and
- (iv) Hundred bilingual speakers of SMG and a second language (Group IV<sub>BI</sub>; participants' second language is shown in Table 3).

Group IV<sub>BI</sub> is further divided in three subgroups:

- (i) Thirty-three simultaneous bilinguals that were exposed to both their languages from birth and grew up mainly in Greece;
- (ii) Twenty-five heritage speakers that were educated exclusively abroad and learned SMG mainly through the home environment; and

- (iii) Forty-two L1 attriters who grew up in Greece as monolingual speakers of SMG and got exposed to their second language only as adults.

Following recent research (Kaltsa et al., 2015), participants in this last subgroup have spent at least 7 years abroad at the time of testing in order to ensure adequate exposure to the other language. Table 3 presents the type distribution of bilinguals within Group IV<sub>BI</sub> and shows the demographics of participants in this group.

All participants were literate adults that had completed secondary education in (mostly public) mainstream schools and were asked to report whether they had a history of neurological or behavioral problems as well as whether they received any speech-pathology treatment. Exclusion criteria included absence of normal articulation, hearing, and (corrected-to-) normal vision, neurological or behavioral problems, and language delay, based on participants' self-report. Participants that reported receiving speech-pathology treatment, a history of neurological and/or behavioral problems and use of hearing aid were excluded from the analyzed results. All bilingual participants have stated that one of their native languages is SMG. Bilectal participants were born and educated in Cyprus, but due to sociolinguistic reasons, they varied in stating that their native language is CG, SMG, both, or simply Greek. All participants from Group I<sub>HG</sub> and Group IV<sub>BI</sub> were tested through an online platform (LimeSurvey<sup>4</sup>), while some participants of Group II<sub>GC</sub> and Group III<sub>GC-GR</sub> were tested in our lab.

An important, final note is necessary with respect to the linguistic identity of the bilingual participants. This study is about the perception of *native* grammatical variants. Rothman and Treffers-Daller (2014) argued that monolingualism and nativeness are often used synonymously in an exclusive way. We take their lead in assuming that bilingual speakers, including heritage language learners, are native speakers too. According to Rothman (2009: p. 156), "a language qualifies as a heritage language if it is a language spoken at home or otherwise readily available to young children, and crucially this language is not a dominant language of the larger (national) society." In the context of the present study, heritage speakers are of interest because it has been argued that their performance may differ from that of non-heritage speakers of the same language with

**TABLE 3 | Languages and types of bilinguals.**

| Languages            | L1 Attriters | Simultaneous bilinguals | Heritage speakers | Total |
|----------------------|--------------|-------------------------|-------------------|-------|
| Albanian             |              | 4                       |                   | 4     |
| Bulgarian            |              | 1                       | 1                 | 2     |
| Danish               | 5            | 2                       | 1                 | 8     |
| English              | 3            | 6                       | 3                 | 12    |
| French               | 1            | 2                       | 2                 | 5     |
| German               | 4            | 7                       | 6                 | 17    |
| German and Norwegian |              | 1                       |                   | 1     |
| Hungarian            | 2            |                         |                   | 2     |
| Norwegian            | 12           | 3                       | 1                 | 16    |
| Romanian             |              | 1                       |                   | 1     |
| Russian              |              | 2                       |                   | 2     |
| Spanish              | 1            |                         |                   | 1     |
| Swedish              | 14           | 3                       | 11                | 28    |
| Turkish              |              | 1                       |                   | 1     |
| Total                | 42           | 33                      | 25                | 100   |

<sup>4</sup>LimeSurvey is an open-source web application used to develop, publish, and collect responses to on-line and off-line surveys.

respect to the amount of variation attested (e.g., Lohndal and Westergaard, 2016; see also Montrul, 2002, 2008). All in all, we consider all our speakers, monolinguals, bilinguals, and billectals, as native speakers of their respective language(s), based on their self-report.

## MATERIALS AND METHODS

This study employed the written variety-judgment task used in Leivada et al. (forthcoming). This task contains a total of 45 sentences, 30 of which involve the presence of morphemes, syntactic structures, graphemes corresponding to phonological variants, or lexical items that are CG-specific. Each of the 30 sentences includes only one dialectal element. Fifteen sentences function as fillers; these are acceptable sentences of SMG with no dialectal element present. In order to exclude random performance in any linguistic area or condition, each area of testing (syntax, morphology, semantics, phonology, and the lexicon) involves two conditions (e.g., two types of morphemes, graphemes, etc.), and each condition has three items of the attested variant (see **Table 4**). All four core levels of linguistic analysis are examined, plus the lexicon, by the same number of test sentences for each ( $n = 6$ ). All groups of test sentences were randomized across conditions.

The five areas and their conditions are the following:

- (A) Syntax was tested by (i) clitic placement that is licit in CG but not in SMG (i.e., CG enclisis in a declarative clause in indicative mood vs. SMG proclisis in the same syntactic environment) and (ii) the realization of syntactic case that is licit in CG but not in SMG (e.g., *opos* 'like' + accusative in CG vs. *opos* 'like' + nominative in SMG).
- (B) Morphology was investigated through (i) the CG-specific diminutive suffix *-ui* and (ii) the CG marking of grammatical aspect, which in some cases (one of which is the verb *katalavo* 'understand', used in the experiment) appears in a morphological combination that is unavailable in SMG, namely, perfective aspect with present tense; in SMG, present tense can only take the morphological suffix of imperfectivity.
- (C) Semantics was examined by (i) the use of nouns ( $n = 3$ ) that exist in both SMG and CG, but with a different meaning across varieties (e.g., *pis:a* means 'chewing gum' in CG, but 'tar' in SMG; in a context where only the former meaning is allowed, the use of this word should be marked as dialectal); and (ii) the use of verbs ( $n = 3$ ) that have different meanings or different thematic requirements across the two varieties.
- (D) Phonology was assessed through nouns that involve a phoneme that is CG-specific (e.g., aspirated stops). Apart from the phoneme in question, these nouns are used identically in both varieties. Given that CG lacks an official writing system, the items were written as they appear in dictionaries and thesauri.
- (E) Lexicon was tested through the use of CG-specific lexical items: verbs ( $n = 3$ ) and nouns ( $n = 3$ ).

**TABLE 4 | List of areas and conditions tested.**

| Area [six items per area] | Condition [three items per condition]  |
|---------------------------|----------------------------------------|
| Syntax                    | Clitics<br>Case                        |
| Morphology                | Diminutive suffix <i>-ui</i><br>Aspect |
| Semantics                 | Nouns<br>Verbs                         |
| Lexicon                   | Nouns<br>Verbs                         |
| Phonology                 | tʃ/f/dʒ<br>kʰ/tʰ/pʰ                    |

Participants completing the task online were presented with each sentence in written form and were asked to read through each sentence carefully and to classify it as either SMG or CG/dialect (Group II<sub>GC</sub> and Group III<sub>GC-GR</sub> were given the former option and all the other groups the latter). Sentences were presented in lower case letters, one at a time, and the software did not allow participants to go back and change their answer. Also, they did not have the option to skip a question. Instructions, written in SMG, were displayed at the top of the window at all times.

Those participants that completed the task in our lab were given a list of sentences of the same format in the form of a booklet, with material presented as in the online task. They were given the same instructions and were 'supervised' by a researcher in order to avoid any self-corrections. All participants had no information with respect to how many sentences involved dialectal elements or how many dialectal parts were present per sentence. It was explained to all participants that the presence of even a single dialectal element sufficed to render the classification of a sentence as CG/dialect. Participants had no time limits. Overall, the task took no more than 20 min to complete. This study was carried out in accordance with the recommendations of the Cyprus National Bioethics Committee, with written informed consent from all subjects.

## RESULTS

A dataset of 16,245 sentences was analyzed. We measured four types of responses:

- (i) *Correct responses to test items* — identifying a test item as having a dialectal element and correctly classifying the test item as CG/dialect.
- (ii) *Wrong responses to test items* — identifying no dialectal element in a test item that has one and failing to classify the test item as CG/dialect.
- (iii) *Correct responses to fillers* — identifying no dialectal element in a filler and correctly classifying the test item as SMG.
- (iv) *Over-corrections* — identifying a filler as having a dialectal element when there is none and failing to classify the test item as SMG.

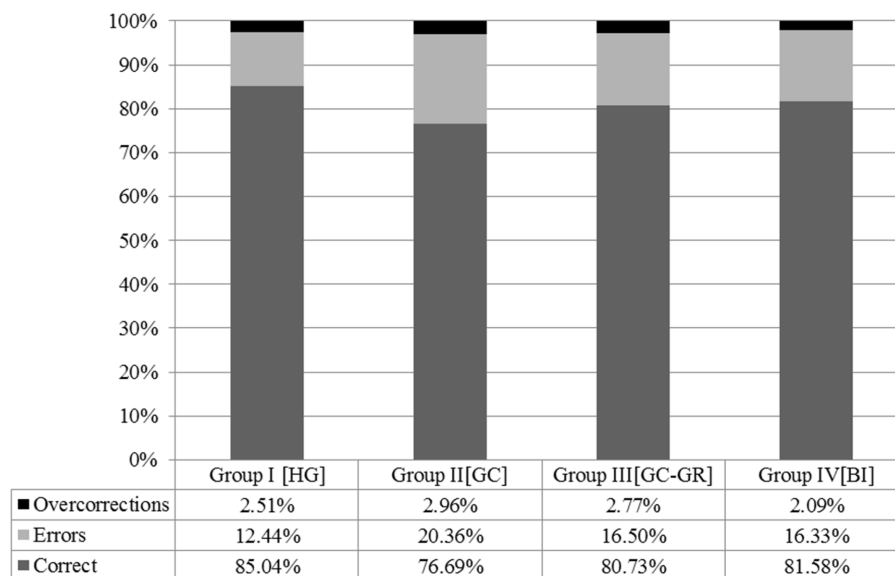


FIGURE 1 | Overall performance across groups.

**Figure 1** presents the overall performance across groups. Correct responses for both test items and fillers are grouped together. Wrong responses in test items are presented as ‘errors’ and over-corrections are shown separately. According to a univariate ANOVA test, all types of bilinguals performed similarly in terms of overall errors [ $F_{(1,97)} = 0.162, p = 0.85$ ], hence they are grouped together in the presentation of the results.

**Figure 1** shows that Group I<sub>HG</sub> (the monolinguals) performed significantly better than Group II<sub>GC</sub> (the bilectals that had not lived in Greece). A univariate ANOVA test showed that there are statistically significant differences between the four groups presented in **Table 2** [ $F_{(3,357)} = 23.61, p < 0.05$ ]. A *post hoc* Tukey analysis showed that the differences in terms of errors, including over-corrections, are statistically significant across all groups, with the exception of the difference between Group III<sub>GC-GR</sub> and Group IV<sub>BI</sub> ( $t = 0.74, Pr = 0.88$ ).

Errors were then analyzed across all conditions. **Figure 2** shows overall percentage of errors in each area of testing: syntax, semantics, morphology, phonology and the lexicon. Fillers are not presented in **Figure 2**; they are analyzed separately below.

## Syntax

Syntax proved particularly difficult for the Groups II<sub>GC</sub> and III<sub>GC-GR</sub>, with the former performing almost at chance level. As **Figure 2** shows, both bilectal groups performed double the amount of errors of their monolingual peers. Bilinguals (Group IV<sub>BI</sub>) performed worse than monolinguals (Group I<sub>HG</sub>), but considerably better than bilectals (Group II<sub>GC</sub>); the differences are statistically significant in both cases according to a Tukey analysis [ $(t = -3.21, Pr = 0.007)$  and  $(t = 5.78, Pr < 0.001)$  respectively].

## Semantics

Semantics is another domain where Group II<sub>GC</sub> stands out as the only group with errors above 40%. The other group of bilectals performed better (Group III<sub>GC-GR</sub>), but the differences are not so pronounced in semantics as were in the case of syntax. According to a Tukey analysis, the differences that reach statistical significance are found between the bilectals of Group II<sub>GC</sub> and both bilinguals ( $t = 4.31, Pr < 0.001$ ) and monolinguals ( $t = -5.41, Pr < 0.001$ ).

## Morphology

Compared to the syntactic and semantic conditions, participants did better in morphology. Once more, the highest error rate was at 15.8% for Group II<sub>GC</sub> with all other error rates below 10%. The only differences that are statistically significant are between the bilectals of Group II<sub>GC</sub> and all other groups (Group II<sub>GC</sub>-Group I<sub>HG</sub>:  $t = -5.12, Pr < 0.001$ ; Group II<sub>GC</sub>-Group III<sub>GC-GR</sub>:  $t = 2.69, Pr = 0.03$ ; Group II<sub>GC</sub>-Group IV<sub>BI</sub>:  $t = 5.28, Pr < 0.001$ ). For the first time, the group that performed best is not that of monolinguals; bilinguals have an even lower percentage of errors, although the difference between the two is not statistically significant based on a Tukey analysis ( $t = 0.15, Pr = 0.99$ ).

## Lexicon

Concerning the lexical condition, Group II<sub>GC</sub> has again the highest percentage of errors at 18.5%, followed by the bilinguals (Group IV<sub>BI</sub>). The following differences across groups are statistically significant according to a Tukey analysis: Group I<sub>HG</sub>-Group II<sub>GC</sub> ( $t = -5.61, Pr < 0.001$ ), Group II<sub>GC</sub>-Group III<sub>GC-GR</sub> ( $t = 3.21, Pr = 0.007$ ), and Group I<sub>HG</sub>-Group IV<sub>BI</sub> ( $t = -3.17, Pr = 0.008$ ).

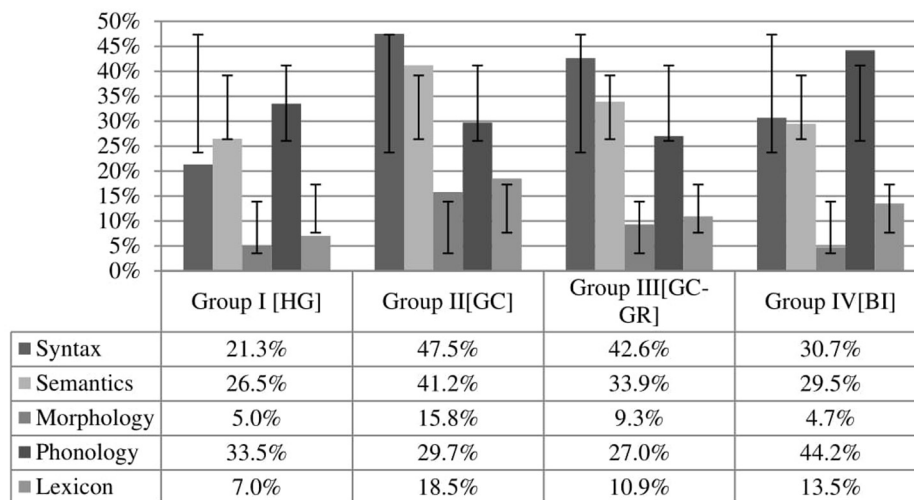


FIGURE 2 | Errors across groups in the five areas of testing.

## Phonology

Figure 2 shows that there is a clear pattern of performance across groups in all conditions apart from phonology, where both monolinguals and bilinguals performed worse than bilectals. Statistically significant differences are found when comparing the bilinguals with any of the other three groups (Group IV<sub>BI</sub>-Group I<sub>HG</sub>:  $t = -3.93$ ,  $Pr < 0.001$ ; Group IV<sub>BI</sub>-Group II<sub>GC</sub>:  $t = -5.34$ ,  $Pr < 0.001$ ; Group IV<sub>BI</sub>-Group III<sub>GC-GR</sub>:  $t = -5.49$ ,  $Pr < 0.001$ ).

The reason for this performance has to do with the task at hand: Phonology was tested through orthography, which may not be the ideal vehicle for testing this domain. As noted in Leivada et al. (forthcoming), the presence of aspirated stops in the task—which are frequently used in CG but do not exist in SMG—is represented in written form through a double occurrence of the relevant consonant. For example, the word *pit:a* ‘pie’ involves one  $\tau$  ‘t’ in SMG, but two in CG. Given that in previous forms of Greek, this word was spelled with  $\tau\tau$ , one can hypothesize that participants of the monolingual group might be familiar with this form, thus failing to mark the relevant sentence as dialectal. Bilectals, however, are strongly aware of this phonological discrepancy since it is one of the most salient characteristics of CG, hence they mark the relevant test structure as such (Leivada et al., forthcoming). Also, in the present study, some of the bilingual participants from Group IV<sub>BI</sub> had been educated mainly in another language (see the subgroups in Table 2) and may not have been able to identify orthographical mismatches from the correct form.

For this very reason, we re-did the overall error analysis without phonology. Not taking phonology into account, a Tukey analysis shows that error differences across all groups remain statistically significant, with the exception of the difference between Group I<sub>HG</sub> and Group IV<sub>BI</sub> ( $t = -2.42$ ,  $Pr = 0.07$ ). In other words, if phonology is disregarded, *monolinguals and bilinguals pattern together* and behave differently than both groups of bilectals.

## Fillers

Fillers reveal an interesting finding of the present study. Recall that fillers are sentences that do not involve any dialectal element. Therefore, mistakes in fillers can be viewed as over-corrections: Participants identify an element as dialectal where there is none. Figure 3 shows such over-corrections across groups, together with the total percentage of errors with and without phonology.

Running a Tukey analysis of over-corrections across groups,<sup>5</sup> statistically significant differences are found only between the bilectals of Group II<sub>GC</sub> and both bilinguals ( $t = 4.94$ ,  $Pr < 0.001$ ) and monolinguals ( $t = -4.56$ ,  $Pr < 0.001$ ).

A comparative analysis of errors excluding the outlier does not change the aforementioned result that monolinguals and bilinguals pattern together, if phonology is not taken into account. A Tukey analysis confirmed that even if the outlier is excluded, the differences between all groups remain significant apart from the difference between Group I<sub>HG</sub> and Group IV<sub>BI</sub> ( $t = -2.203$ ,  $Pr = 0.12$ ). In other words, the exclusion or inclusion of the outlier does not alter one of the most crucial findings of this experiment: If phonology is disregarded, as it should, monolinguals and bilinguals behave alike. Figures 4 and 5 show patterns of errors across conditions, excluding the outlier.

An analysis of possible factors affecting Groups’ performance was also performed and revealed that gender and place and/or format of the implementation of the task did not affect the tendencies and results observed above. A univariate ANOVA showed that the bilectals of Group II<sub>GC</sub> that took the study in the lab ( $n = 39/100$ ) did not perform differently in a statistically

<sup>5</sup>One filler was excluded from these analyses because it involves the word *riya* ‘ruler,’ which, despite being listed in SMG dictionaries with this meaning, is not frequently used in the language (another word is used instead)—but it is frequently used in CG. As a result, some participants in all groups classified this test sentence as ‘dialectal’ (Group I<sub>HG</sub> at 85%, Group II<sub>GC</sub> at 44%, Group III<sub>GC-GR</sub> at 65% and Group IV<sub>BI</sub> at 71%), whereas in the test design it counts as filler (i.e., non-dialectal/SMG). Figure 3 includes this filler and Figures 4 and 5 exclude it.

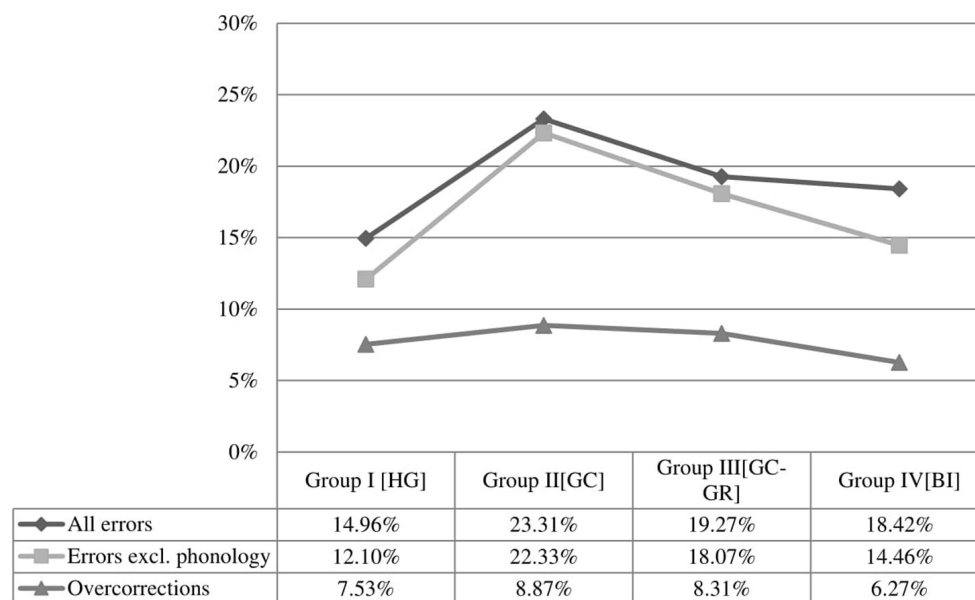


FIGURE 3 | Overall errors with and without phonology, and over-corrections.

significant way from the ones that completed the experiment online [ $F_{(1,97)} = 2.70, p = 0.10$ ]. With respect to the bilectals of Group III<sub>GC-GR</sub>, a univariate ANOVA showed that the participants that took the study in the lab ( $n = 11/61$ ) performed differently in a statistically significant way from the ones that completed the experiment online [ $F_{(1,59)} = 5.125, p = 0.02$ ], with the latter being better (mean errors 8.28 vs. 10.45 in the lab-based participants). The low number of lab-based participants should be taken into account in interpreting the difference in the performance of these two groups.

## DISCUSSION

The findings of this study suggest that exposure to two different grammars affects speakers' performance in variety judgment tasks in some levels of linguistic analysis. This result agrees with the differential performance across monolinguals and bilinguals that has been reported in the literature (Jacobson and Cairns, 2008). The relevant literature refers to the performance of bilingual populations, covering *languages* rather than different varieties of the same language. It is less clear what happens when the developmental trajectory features varieties of one language rather than (adequately) different languages.

It has been argued earlier that since bilectals have exposure to both varieties, it is expected that they would perform better than all the other groups in classifying correctly the test stimuli, precisely because they are familiar both with the Standard and the dialectal element. Figures 3 and 4 suggest that this is not the case; overall, bilectals were less accurate than all the other groups. This could be due to different factors. One possibility is that non-standardization blurs the boundaries of grammatical variants, hence bilectals become less accurate in identifying the variety to

which a grammatical variant pertains. Under another scenario, the differential performance of bilectals derives from the fact that these speakers are competent in two linguistic systems. Figure 3 shows that the group of bilinguals (Group IV<sub>BI</sub>) performed better than both groups of bilectals, and this should not happen if the second scenario was on the right track, because bilinguals have exposure to two linguistic systems too, yet they perform better than bilectals. This entails that the differential performance of bilectals is due to other factors.

For instance, linguistic proximity, defined here as the typological closeness between the varieties one is exposed to, is an important factor that characterizes language development in bilectal settings (Grohmann, 2014; Grohmann and Kambanaros, 2016). Recent research in speakers of closely related varieties has revealed that factors such as non-standardization and close proximity affect speakers' perception of the variants that exist in their native repertoires in a way that is absent from the performance of their monolingual peers (Leivada et al., forthcoming). The present study is the first to tackle the question of whether this differential performance is the result of being exposed to more than one linguistic system in the course of development or of factors inherent to the dialect design (e.g., lack of standardization, unclear boundaries between variants, linguistic proximity between the varieties, awareness of the fact that some of the structures of the L-variety might be considered incorrect by speakers of the standard, etc.). Pursuing this novel comparison between monolinguals, bilinguals and bilectals, we uncovered significant differences between bilectal and bilingual speakers, providing support for the second scenario: *Bilinguals performed considerably better than bilectals in identifying correctly dialectal elements.*

We aimed to ascertain whether this happens because of the dialect design alone or also because what counts as standard in

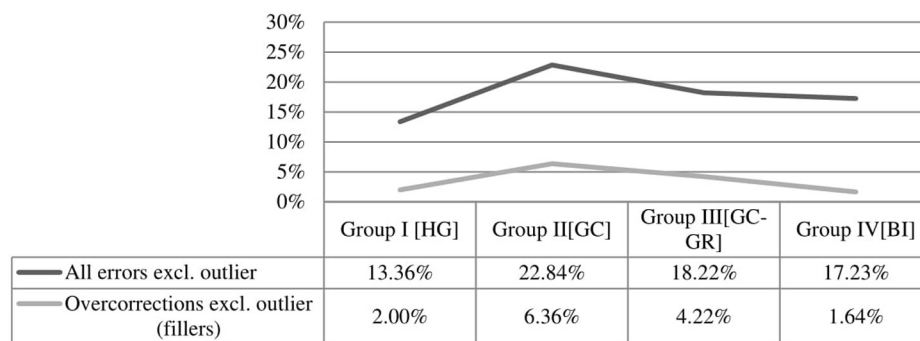


FIGURE 4 | Overall errors without outlier across all conditions and in fillers.

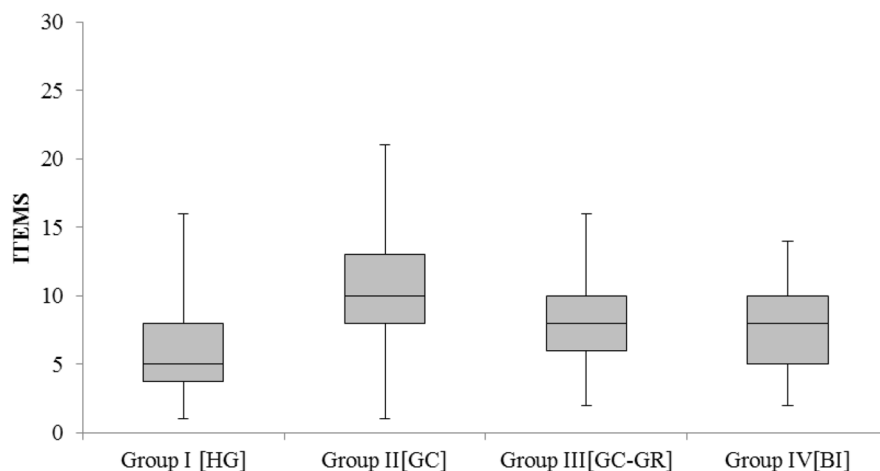


FIGURE 5 | Dispersion of errors without outlier across all conditions and in fillers.

Cyprus may not always correspond to SMG as spoken in Greece by Hellenic Greeks but to another form of standard (perhaps Cypriot Standard Greek; Arvaniti, 2010). For this purpose, we included a group of participants that grew up as bilectals but spent some time in Greece (>12 months) as adults: Group III<sub>GC-GR</sub>. Results suggest that in the overall calculation of errors, with and without the outlier, Group III<sub>GC-GR</sub> performs more like the bilinguals of Group IV<sub>BI</sub> and less like their bilectal peers that had not lived in Greece (Group II<sub>GC</sub>). At the same time, the bilectals of Group III<sub>GC-GR</sub> were less accurate than the bilinguals of Group IV<sub>BI</sub>, as shown in **Figure 4**. The difference is also evidenced in over-corrections, which can be a marker of linguistic insecurity: As **Figure 4** shows, Group III<sub>GC-GR</sub> performed more than twice the errors of the bilinguals in fillers. The conclusion to be drawn is that exposure to the standard in Cyprus may not always amount to SMG, which is why prolonged exposure to SMG in Greece makes bilectals behave more like true bilinguals. At the same time, factors inherent to the dialect design have influenced the linguistic development of bilectals: Their performance remains less accurate than that of bilinguals, and an increased degree of linguistic insecurity is still manifested in their performance even *after* prolonged exposure to SMG.

Errors in fillers are interesting because they do not amount to missing a dialectal element superimposed on an otherwise standard form, but to identifying a dialectal element where a dialectal element is not present. The attested higher degree of over-corrections in the two bilectal groups cannot be the result of having two grammars. Bilinguals have two grammars too, yet their performance is comparable to that of monolinguals. Therefore, another explanation should be found regarding the higher degree of over-corrections in bilectals. A possible reason is the linguistic insecurity that often characterizes dialect speakers (Toribio, 2000). Bilectal speakers are eager to show that they are competent in the H-variety, and this may be the cause of the higher degree of over-corrections in their performance. This finding is in perfect agreement with data from child language that come from the bilectal context of Cyprus. As Grohmann and Leivada (2012) have argued, bilectal children's process of building a sociolinguistic repertoire primarily involves the need to resolve linguistic anxiety and adjust to the H-variety. It is likely that this factor of dialect design drives linguistic performance of bilectal speakers even well past the acquisition period (the 'socio-syntax of development hypothesis'). This is true also of the group of bilectals that have lived in Greece (Group III<sub>GC-GR</sub>). Despite

their prolonged exposure to SMG, they show a higher degree of over-corrections compared to their monolingual and bilingual peers.

A last result that is worth highlighting relates to the performance of the three subgroups within Group IV<sub>BI</sub>. Their performance was found to be so similar that they were grouped together in all calculations mentioned above. This finding agrees with the results of Kaltsa et al. (2015) who examined monolingual speakers of SMG and two types of bilingual speakers (heritage speakers and L1 attriters) of SMG and Swedish in a sentence-picture matching decision task. They found differences in anaphora resolution of overt and null subject pronouns between monolinguals and bilinguals, but *not* between the two groups of bilinguals. The only difference between the two bilingual groups was found in reaction times, with heritage speakers being faster than L1 attriters. The absence of a difference in the off-line measure led Kaltsa et al. (2015, p. 266) to conclude that their results “do not support an age of onset or differential input effects on bilingual performance in pronoun resolution.” Our results seem to fully support this conclusion too across different grammatical conditions and domains of linguistic analysis.

## CONCLUSION

The results of the present study have confirmed the findings of previous research showing that bilinguals perform differently than monolinguals in acceptability and variety judgment tasks *only* in some linguistic domains. In the written variety-judgment task employed here, if phonology is not taken into account, the differences between monolinguals and bilinguals are not statistically significant. In addition, it was found that bilectals performed worse than both bilinguals and monolinguals, and that the defining characteristics of bilectalism and the dialect design (e.g., linguistic proximity and grammatical fluidity) affect speakers' performance.

The notion of linguistic proximity is important for the interpretation of the results of the bilectals in the following way. Even though they were the ones that had exposure to both varieties, thus possibly being in a more privileged position to

correctly identify the test stimuli compared to the monolinguals, they turned out to be less accurate than the monolinguals. Proximity plays a role in that it facilitates the emergence of mesolectal varieties that blur the limits of different lects. This in turn makes the bilectals less accurate in distinguishing the lect to which each grammatical variant pertains.

This study has shown that the linguistic insecurity that is often found in bilectal speech communities (Toribio, 2000) persists in the form of over-corrections even after prolonged exposure to the H-variety. Last, our comparison of three groups of bilingual speakers did not show significant differences between them, granting new support to the argument of Kaltsa et al. (2015) that age of onset and differential input do not affect performance in off-line measures.

## ETHICS STATEMENT

According to the local legislation in Cyprus, ethical approval was not required for this type of study. The study was, however, conducted in accordance with the declaration of Helsinki and written informed consent was obtained from each participant.

## AUTHOR CONTRIBUTIONS

EL designed the experiment in consultation with EP, MK, and KKG; EL and EP recruited the participants; EL and EP analyzed the results; EL, EP, MK, and KKG wrote up the paper.

## ACKNOWLEDGMENTS

We would like to thank all the participants who took place in this study. We also thank Adrian Monroy and Nikoleta Christou for their help with the statistical analyses. We are grateful to the people and associations that helped us recruit Greek-speaking populations in Scandinavia, especially the Greek School of Oslo, Paraskevi Karapostoli, Anastasia Teterina, Kyriaki Papadopoulou Samuelsen, and Andreas Bouras.

## REFERENCES

- Arvaniti, A. (2010). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *Mediterr. Lang. Rev.* 17, 15–45.
- Bard, E. G., Robertson, D., and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language* 72, 32–68.
- Bialystok, E., Craik, F. I. M., Klein, R., and Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychol. Aging* 19, 290–303.
- Cheshire, J., and Stein, D. (1997). *Taming the Vernacular: From Dialect to Written Language*. Harlow: Longman.
- Cornips, L. (2006). “Intermediate syntactic variants in a dialect-standard speech repertoire and relative acceptability,” in *Gradi-Ence in Grammar: Generative Perspectives*, eds G. Fanselow, C. Féry, R. Vogel, and M. Schlesewsky (Oxford: Oxford University Press), 85–105.
- Cornips, L., and Poletto, C. (2005). On standardising syntactic elicitation techniques (part 1). *Lingua* 115, 939–957.
- Foursha, C. D., Austin, J. B., and Van de Walle, G. A. (2006). “Is language processing identical in monolinguals and early, balanced bilinguals?” in *Proceedings of the Supplement of the 30th Boston University Conference on Language Development*, eds D. Bamman, T. Magnitskaia, and C. Zaller, Boston, MA.
- Garraffa, M., Beveridge, M., and Sorace, A. (2015). Linguistic and cognitive skills in Sardinian-Italian bilingual children. *Front. Psychol.* 6:1898. doi: 10.3389/fpsyg.2015.01898
- Grohmann, K. K. (2014). Towards comparative bilingualism. *Linguist. Approaches Biling.* 4, 336–341.
- Grohmann, K. K., and Kambanaros, M. (2016). The gradience of multilingualism in typical and impaired language development: positioning bilectalism within comparative bilingualism. *Front. Psychol.* 7:37. doi: 10.3389/fpsyg.2016.00037
- Grohmann, K. K., and Leivada, E. (2012). “Interface ingredients of dialect design: Bi-x, socio-syntax of development, and the grammar of cypriot Greek,” in *Towards a Bilingual Understanding of Grammar: Essays on Interfaces*, ed. A. M. Di Sciullo (Amsterdam: John Benjamins), 239–262.

- Henry, A. (2005). Non-standard dialects and linguistic data. *Lingua* 115, 1599–1617.
- Ioannidou, E. (2002). *This Ain't my Real Language, Miss: On Language and Ethnic Identity among Greek Cypriot Students*. Ph.D. dissertation, University of Southampton, Southampton.
- Jacobson, P. F., and Cairns, H. S. (2008). Considering linguistic input in a bilingual situation: implications for acquisition. *Lang. Speech Hear. Serv. Sch.* 39, 352–364. doi: 10.1044/0161-1461(2008/033)
- Kaltsa, M., Tsimpli, I.-M., and Rothman, J. (2015). Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: evidence from pronominal resolution. *Lingua* 164, 266–288.
- Karyolemu, M. (2006). "Reproduction and innovation of communicative patterns in a former 'diglossic' community," in *Reproduktionen und Innovationen in Sprache und Kommunikation verschiedener Sprachkulturen*, ed. R. Muhr (Berlin: Peter Lang), 39–56.
- Keller, F. (2000). *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D. dissertation, University of Edinburgh, Edinburgh.
- Laka, I. (2012). "More than one language in the brain," in *Language, from a Biological Point of View: Current Issues in Biolinguistics*, eds C. Boeckx, M. Horno-Chéliz, and J.-L. Mendivil-Giró (Newcastle: Cambridge Scholars Publishing), 184–207.
- Leivada, E., Kambanaros, M., Taxitari, L., and Grohmann, K. K. (forthcoming). (Meta)linguistic abilities of bilingual educators: the case of Cyprus. *Int. J. Biling. Educ. Biling.*
- Lohndal, T., and Westergaard, M. (2016). Grammatical gender in american Norwegian heritage language: stability or attrition? *Front. Psychol.* 7:344. doi: 10.3389/fpsyg.2016.00344
- Milroy, J. (2001). Language ideologies and the consequences of standardization. *J. Sociolinguist.* 5, 530–555.
- Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism* 5, 39–68.
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: John Benjamins.
- Moreno, S., Bialystok, E., Wodniecka, Z., and Alain, C. (2010). Conflict resolution in sentence processing by bilinguals. *J. Neurolinguist.* 23, 564–579.
- Papadopoulou, E., Leivada, E., and Pavlou, N. (2014). Acceptability judgments in bilingual populations: competition, gradience and socio-syntax. *Linguist. Var.* 14, 109–128.
- Papapavlou, A., and Yiakoumetti, A. (2003). "The analysis and classification of non-standard occurrences in the writing of Greek Cypriot primary-school children," in *Selected Papers on Theoretical and Applied Linguistics*, ed. E. Mela-Athanasopoulou (Thessaloniki: Aristotle University of Thessaloniki), 340–348.
- Pavlou, P., and Christodoulou, N. (2001). Bidialectism in Cyprus and its impact on the teaching of Greek as a foreign language. *Int. J. Appl. Linguist.* 11, 75–91.
- Pavlou, P., and Papapavlou, A. (2004). Issues of dialect use in education from the Greek Cypriot perspective. *Int. J. Appl. Linguist.* 14, 243–258.
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: romance languages as heritage languages. *Int. J. Biling.* 13, 155–163.
- Rothman, J., and Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Appl. Linguist.* 35, 93–98.
- Rowe, C., and Grohmann, K. K. (2013). Discrete bilingualism: towards co-overt prestige and diglossic shift in Cyprus. *Int. J. Sociol. Lang.* 224, 119–142.
- Rowe, C., and Grohmann, K. K. (2014). Canaries in a coal mine: native speakerhood and other factors as predictors of moribundity, death, and diglossic shift in Cypriot Greek. *Mediterr. Lang. Rev.* 21, 121–142.
- Schütze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago, IL: University of Chicago Press.
- Sophocleous, A., and Wilks, C. (2010). Standard Modern Greek and Greek-Cypriot dialect in kindergarten classroom interaction: teachers' and learners' language attitudes and language use. *Lang. Cult. Curriculum* 23, 51–69.
- Sprouse, J., and Almeida, D. (2013). "The role of experimental syntax in an integrated cognitive science of language," in *The Cambridge Handbook of Biolinguistics*, eds C. Boeckx and K. K. Grohmann (Cambridge: Cambridge University Press), 181–202.
- Sprouse, J., Schütze, C., and Almeida, D. (2013). Assessing the reliability of journal data in syntax: linguistic Inquiry 2001–2010. *Lingua* 134, 219–248.
- Toribio, A. J. (2000). Language variation and the linguistic enactment of identity among Dominicans. *Linguistics* 38, 1133–1159.
- Westergaard, M., Mitrofanova, N., Mykhaylyk, R., and Rodina, Y. (2016). Crosslinguistic influence in the acquisition of a third language: the linguistic proximity model. *Int. J. Biling.* 1–17. doi: 10.1177/1367006916648859

**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 © 2017 Leivada, Papadopoulou, Kambanaros and Grohmann. 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.



# Acquisition of Classifier Constructions in HKSL by Bimodal Bilingual Deaf Children of Hearing Parents

Gladys W. L. Tang\* and Jia Li

Department of Linguistics and Modern Languages, Centre for Sign Linguistics and Deaf Studies, The Chinese University of Hong Kong, Shatin, Hong Kong

## OPEN ACCESS

### Edited by:

Gary Morgan,  
City University of London,  
United Kingdom

### Reviewed by:

Kadir Gokgoz,  
Bogaziçi University, Turkey  
Deborah Chen Pichler,  
Gallaudet University, United States

### \*Correspondence:

Gladys W. L. Tang  
gtang@cuhk.edu.hk

### Specialty section:

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

**Received:** 17 July 2017

**Accepted:** 14 June 2018

**Published:** 23 July 2018

### Citation:

Tang GWL and Li J (2018) Acquisition  
of Classifier Constructions in HKSL by  
Bimodal Bilingual Deaf Children of  
Hearing Parents.  
Front. Psychol. 9:1148.  
doi: 10.3389/fpsyg.2018.01148

The current study focuses on the acquisition of classifier constructions in Hong Kong Sign Language (HKSL) by a group of Deaf children of hearing parents, aided or implanted. These children have been mainstreamed together since kindergarten; but their learning environment supports dual language input in Cantonese and HKSL on a daily basis. Classifier constructions were chosen because previous research suggested full mastery at a late age when compared with other verb types, due to their morphosyntactic complexity. Also, crosslinguistic comparison between HKSL and Cantonese reveals differences in verb morphology as well as word order of the structures under investigation. We predicted that verb root and word order were the two domains for crosslinguistic interaction to occur. At the general level, given the specific learning environment and dual input condition, we examined if these Deaf child learners could ultimately acquire classifier constructions. Fifteen Deaf children divided into four groups based on duration of exposure to HKSL participated in the study. Two Deaf children born to Deaf parents and three native HKSL signers served as controls. A picture description task was designed to elicit classifier constructions containing either a transitive, a locative existential or a motion directional predicate. The findings revealed Deaf children's gradual convergence on the adult grammar despite late exposure to HKSL. Evidence of crosslinguistic influence on word order came from the Deaf children's initial adoption of a Cantonese structure for locative existential and motion directional predicates. There was also a prolonged period of adherence to the SVO order across all grades. However, within this SVO structure, the verb revealed increasing morphological complexity as a function of longer duration of exposure. We interpreted the findings using Language Synthesis, arguing that it was the selection of morphosyntactic features in Numeration that triggered crosslinguistic interaction between Cantonese and HKSL with bimodal bilinguals.

**Keywords:** bimodal bilingualism, word order, classifier constructions, language acquisition, HKSL, Cantonese, deaf children, coenrollment

## INTRODUCTION

How deaf and hard-of-hearing children acquire language has always attracted attention among researchers in linguistics, speech and language pathology, and deaf education. In recent years, due to advancement in hearing technology, one also saw an increasing number of signing Deaf children demonstrating knowledge of spoken language either through print, or print and speech. To appreciate this development, one needs to understand the demography of Deaf children. Generally speaking, Deaf children who are born to Deaf parents (i.e., DDs) may acquire sign language since birth, and spoken language when they begin to receive speech training which usually comes as early as if not earlier than 1 year old. Hearing children who are born to Deaf parents (i.e., Kodas) usually acquire sign language and spoken language much earlier in life, if not simultaneously at birth. A great majority of Deaf children are born to hearing parents (i.e., DHs), and whose exposure to sign language depends largely on the type of formal education they receive. Some start to receive sign language exposure when their parents enroll them into schools for the deaf at various ages. Although integrative/inclusive education nowadays has led to a majority of DHs being mainstreamed without exposure to sign language, there is a small number of them whose education is facilitated by sign interpreters, hearing teachers who can sign and sometimes Deaf teachers. One such mode of bilingual education for the deaf in mainstream education is coenrollment, whereby a critical mass of deaf students study with hearing students in a mainstream classroom, supported by sign language and spoken language. This study focuses on this particular group of signing Deaf children whose parents enroll them into the Sign Bilingualism and Co-enrollment Education Programme in Hong Kong, generally referred to as the “SLCO Programme.” Through naturalistic interactions, these children receive Hong Kong Sign Language (HKSL) input consistently from 7 to 8 Deaf teachers and a critical mass of Deaf peers on a daily basis in the classroom/school setting (see section Participants), in addition to spoken language from their hearing teachers and peers.

Recently, researchers attempt to examine the bilingual acquisition of Kodas and DDs within the framework of *bimodal bilingualism*, defined as acquisition and use of a sign language and a spoken language that stem from the visual-gestural and the auditory-oral modalities respectively. From a child language perspective, bimodal bilingualism has been associated with bilingual first language acquisition, which, in the spoken language literature, is further categorized into *simultaneous* and *sequential bilingual acquisition*. The former refers to acquisition of two languages at the same time since birth while the latter requires exposure to a second language at a very young age and usually no later than 5 (see Meisel, 2011). A general characteristic of bimodal bilingual acquisition is *code blending*, defined as simultaneous and systematic production of sign and speech<sup>1</sup>. A number of studies targeting Kodas and DDs reveal

a prevalence of congruent code blends and the challenge is how to account for the incongruent ones (Petitto et al., 2001; van den Bogaerde and Baker, 2005; Lillo-Martin et al., 2010; Donati and Branchini, 2013; Fung and Tang, 2017). Additionally, in the spoken language literature, while it is generally agreed that bilingual children separate the two grammars from earlier on, systematic crosslinguistic influence is also at play. Hulk and Müller (2000) argue for two conditions for crosslinguistic influence to occur, namely interface between pragmatics and syntax, and structural overlap at the surface level. These two conditions have been subject to investigation in many bilingual acquisition studies. As for bimodal bilingualism, research shows that crosslinguistic influence is observed in structures not predicted by such conditions (Lillo-Martin et al., 2010), and findings for the structures that satisfy these two conditions run counter to predictions (Koulidobrova, 2012, 2016). Recently, Language Synthesis (Koulidobrova, 2012, 2016; Lillo-Martin et al., 2012, 2014, 2016) has been proposed to account for the various language interaction effects observed in bimodal bilingualism (see the next section). The proposal is based on MacSwan’s (2000, 2005) accounts for code-switching, in which he argues for one computational system with separate lexicons and separate Phonetic Forms (PFs) for different languages.

This study focuses on another language pair, HKSL and Cantonese, and investigates how Deaf bimodal bilinguals born to hearing parents acquire classifier constructions in HKSL. This structure was chosen because full mastery has been reported late due to its morphosyntactic complexity [ASL (American Sign Language): Supalla, 1982; Schick, 1990; HKSL: Lam, 2009]. Additionally, while sharing SVO as the basic word order, HKSL and Cantonese differ in verb morphology. Cantonese is said to be poor in inflection; verbs are bare and the basic word order is consistently SVO. On the contrary, HKSL is rich in inflection and the morphosyntactic properties of the verb interact with word order changes. These crosslinguistic differences invite an examination of how bimodal bilingual Deaf children develop knowledge of verb morphology and word order in classifier constructions in HKSL. Additionally, we also explore if DHs can acquire knowledge of such complex constructions as a function of duration of exposure, given the fact that they fail to receive early HKSL input since birth. Last, we examine to what extent Language Synthesis may account for the language interaction effects observed in this study. Evidence supporting Language Synthesis is rather limited, hence further exploration to identify the conditions for language interaction effects to occur is necessary.

The paper is organized as follows. We will first summarize the word order issues that have been documented in bimodal bilingual acquisition of a number of language pairs. Then, we introduce Language Synthesis, recently proposed to account for language interaction effects such as code blending, code switching, as well as crosslinguistic influence and transfer (Lillo-Martin et al., 2016). Based on these discussions, we compare the verb root and word order issues with the relevant constructions between Cantonese and HKSL. We then set out some predictions about how crosslinguistic interaction may occur. The experimental procedure, backgrounds of the DHs and the results are then summarized and discussed. At the end of

<sup>1</sup>Emmorey et al. (2008) observe that bimodal bilinguals code blend much more often than code switch, indicating that lexical suppression is more costly than lexical selection. Further, given the articulatory constraints are removed, it is possible for bimodal bilinguals to actually produce both languages simultaneously.

the paper, we will discuss some caveats of the study and offer suggestions for future research.

## BACKGROUND

### Previous Acquisition Research on Word Order in Sign Languages

In the early literature on ASL acquisition, canonical SVO and derived word orders are observed to emerge at an early age among DDs (Newport and Meier, 1985; Lillo-Martin, 1999; Chen Pichler, 2001). However, Lillo-Martin and Berk (2003) found that the two DHs in their study, who were not exposed to an accessible language like ASL until after age 5, had no problem acquiring the canonical SVO order but seldom attempted derived word orders that reflected grammatical dependencies and erred more when they did so. Reports involving bimodal bilinguals especially Kudas are emerging in recent years. Based on the longitudinal data (ages ranged from 2;00 to 4;00) of two ASL-English and one Brazilian Sign Language (Libras)-Brazilian Portuguese (BP) Kudas, Lillo-Martin et al. (2010:272) observed doubling in the English data (e.g., “sleeping mouse sleeping,” Ben 2;01). Putting forward *Language Synthesis* as an overarching framework of analysis, they argued that the doubling phenomenon may be captured by the choice of a functional element with a [+focus] feature from ASL in the Numeration and late insertion of lexical items in English. In another study, the same team of researchers examined the same Kudas’ (age 1;11–4;05) production of *wh*-questions (Lillo-Martin et al., 2012). According to them, English and BP allow fronted and *in-situ wh*-questions only whereas ASL and Libras’s *wh*-questions allow more syntactic options: fronted, *in-situ*, and doubled (e.g., ASL: WHO JOHN SEE WHO “Who did John see?”). Generally speaking, they observed emergence of *in-situ wh*-questions earlier with bimodal bilinguals than monolinguals of either spoken language. Additionally, while monolingual English and BP child acquirers produced fronted *wh*-questions exclusively, bimodal bilinguals’ *wh*-questions in English were fronted, *in-situ* as well as doubled. What is also interesting is that these doubled *wh*-questions began to retreat from the English of the two Kudas after 2;11. Using an elicited production task on a larger sample of Kudas, the researchers found a much higher rate of production of *wh*-initial questions in ASL by the Kudas than the Deaf controls. Recently, Palmer (2015) compared the acquisition of ASL canonical and non-canonical word orders of four bimodal bilinguals, two Kudas and two implanted DDs whose ages ranged from 1;8 to 3;6. While both the Kudas and DDs produced canonical SV and VO orders as early as 23 months, suggesting an early setting of Spec-Head and Head-Complement parameters, they showed little use of non-canonical OV and VS orders when compared with the Deaf controls as reported in Chen Pichler (2001).

In the HKSL context, few acquisition studies focus on the relation between word order and verb root of classifier constructions. Tang et al. (2007) elicited simultaneous constructions from a group of DHs who studied in a school for the deaf (ages ranged from 6 to 13). They used comic strips to elicit narratives from these participants. The DHs did not

introduce the antecedent before the classifier predicate, nor did they sustain the classifier on the non-dominant hand in space that refers to the direct or indirect object. Lam (2009) in a longitudinal study of a DD acquiring HKSL found that the nominal antecedent is usually not overtly expressed but recoverable from the signing discourse. Although the first token of OV order with a classifier predicate involving one argument emerged at age 2;9.29, very few OV or SOV orders were observed throughout. Instead, the Deaf child produced primarily VO (46.67%) and SVO (33.33%) orders with a classifier predicate during the observation period, which we interpret to be illicit word orders for this structure. Lam (2009) ascribed it to the optionality of object shift that delayed full acquisition. Indeed, one needs to address why the Deaf child accepted an (S)VO order for classifier constructions. We predict that language interaction effect associated with the canonical SVO order of Cantonese and HKSL might be the cause of this acquisition phenomenon (see section Crosslinguistic Comparison and Acquisition Predictions).

### Emerging Accounts for Language Interaction Effects

As previously discussed, *Language Synthesis* has been put forward to account for code switching, code blending, crosslinguistic influence in early language development, transfer in second language acquisition, and calquing in language contact situations (see Lillo-Martin et al., 2016 for a detailed illustration). This model has its basis in Distributed Morphology, which posits that it is the selected roots and atomic features in the Numeration (i.e., List 1) that enter the syntactic computation, and insertion of Vocabulary Items from List 2 is a late phenomenon taking place after Spell-Out to the PF branch (Harley, 2014). According to *Language Synthesis*, List 1 and List 2 are the two places at which interaction between Lx and Ly may occur. When the atomic features are selected from Lx but Vocabulary Insertion draws items from Ly, syntactic synthesis (i.e., crosslinguistic influence, transfer, and calquing) results. Embracing two paths toward PF after Spell-out, one for sign and the other for speech, the process allows simultaneous realization of the possible mix of elements from Lx and Ly, resulting in code blending. Additionally, Lillo-Martin et al. (2016) suggest that the apparent crosslinguistic influence is actually bimodal bilingual effects, meaning that in the constant absence of forced language choice (i.e., inhibition), bimodal bilinguals are “accustomed” to practicing choosing grammatical elements between Lx and Ly during Numeration and Vocabulary Insertion. The so-called crosslinguistic influence is only a reflection of bimodal bilinguals’ capacity for language synthesis.

*Language Synthesis* has attracted a lot of debates about its explanatory adequacy, particularly, for cases where bimodal bilinguals produce two independent strings in diverse word orders simultaneously, as in the code blending between Italian/LIS (Italian Sign Language) or Dutch/NGT (Sign Language of the Netherlands) that involves divergent SVO vs. SOV orders (Baker, 2016; Branchini and Donati, 2016), as the example in (1) shows:

(1)

|          |                       |          |                  |
|----------|-----------------------|----------|------------------|
| Italian: | Non                   | ho       | capito           |
|          | not                   | have.1SG | understand. PTCP |
|          |                       | __NEG    |                  |
| LIS:     | UNDERSTAND            | NOT      |                  |
|          | "I don't understand." |          |                  |

(Branchini and Donati, 2016, example 13).

Instead of one mixed Numeration and late Vocabulary Insertion as what Language Synthesis suggests, Branchini and Donati (2016) argue that bimodal bilinguals have at their disposal two separated monolingual Numerations and two parallel syntactic derivations. They have identified three types of code blendings. The first type (Type 1) has one syntactic representation the derivation of which is based on one Numeration and governed by a single grammar of either LIS or Italian. The output displays all the necessary properties of the language dictating the representation. As bimodal bilinguals are equipped with a double spell-out, lexical retrieval from the "governed" language to derive code blending or fragment insertion in code switching can take place at a late stage, hence it will not affect the grammatical representation of the "dictating" language. As a result, the governed language is impoverished in terms of morphological and phonological properties. The second type (Type 2) involves two strings with independent representations and full-fledged morphological and phonological properties, as in (1) above. This type is often observed when two languages have a rigid word order for functional elements (e.g., the position of negators in Italian is preverbal while in LIS it is postverbal). They argued that such occurrences are due to two parallel Numerations and syntactic derivations. The third type (Type 3) like (2) below have two simultaneous strings that "contribute together to form a unique utterance" (Branchini and Donati, 2016 p.21). Type 3 differs from Type 1 in that both language strings are not impoverished in any sense; it also differs from Type 2 in having one mixed, not two separated Numeration which contributes to a single derivation. In example (2), the subject (i.e., I) is provided by Italian while the predicate (i.e., WIN) by LIS. Only when both language strings are taken into account together will the utterance become complete and meaningful. Based on grammaticality judgment and elicited production data, Branchini and Donati confirmed that all three types are part of the Kodas as well as the adults' grammar, hence not developmental. Additionally, Type 3 is akin to what Language Synthesis stipulates, where merging roots and morphemes from two different languages is possible in the Numeration initially.

(2)

|          |          |
|----------|----------|
| Italian: | io       |
|          | 1SG      |
| LIS:     | WIN      |
|          | "I win." |

(Branchini and Donati, 2016, example 37).

Both Language Synthesis and the proposal by Branchini and Donati (2016) share the assumption that bimodal bilinguals are characterized by co-activation and non-inhibition during bilingual processing. They diverge in the theoretical assumptions about (a) whether there is a list of morphosyntactic features or a Lexicon to store lexical items with pre-assembled features; and (b) whether there is only one mixed Numeration to drive a single derivation or two separate Numerations to drive two parallel syntactic derivations. While these proposals are originally developed to account for bilingual first language acquisition of Kodas, it is possible to extend the analyses to examine language interaction effects in the developing grammars of bimodal bilingual Deaf children from hearing families. So far, the Language Synthesis model has been adopted to account for word order data. The current study aims to extend the analysis to the interaction between word order and morphosyntactic features as involved in classifier constructions. Additionally, we adopt Distributed Morphology in our analysis of classifier constructions in HKSL because we assume it is a "list" of morphosyntactic features, not a Lexicon, that forms the basis for Numeration. However, we are open to Branchini and Donati's (2016) proposal for the possibility of two independent Numerations.

## Word Order

### Verb Morphology and Word Order in HKSL

Similar to other sign languages, word order in HKSL interacts with verb morphology. Verbs in HKSL can be categorized into three types, i.e., plain verbs, agreement verbs, and spatial verbs. Plain verbs such as LIKE and THINK are generally without inflectional morphology; spatial verbs like PUT and TAKE can be modified through movement to the R-loci of location arguments, and agreement verbs like HELP, PUSH, and GIVE associate the R-loci with the subject and/or (indirect) object in terms of person and number. According to Sze (2000), the canonical word order in HKSL is SVO with plain verbs (3a,b) and SOV with agreeing and spatial verbs (3c,d)<sup>2</sup>.

(3a)

|       |                     |      |     |
|-------|---------------------|------|-----|
| HKSL: | FATHER              | LIKE | BOY |
|       | "Daddy likes boys." |      |     |

<sup>2</sup>**Notation Conventions:** Following the conventions in the field, glosses for signs are capitalized (e.g., BOOK); glosses for a single sign are underscored (e.g., TAKE\_A\_PLANE); compound signs are marked with ^ (e.g., WHAT\_MONTH^WHAT\_DATE); nonmanuals and their scopes are marked by a line above the glosses; pointing signs are glossed as IX (e.g., IX-1 = first-person pronominal; IX<sub>a</sub> refers to a locus in space); locations are indicated by subscripted letters "a, b, c," whereas the subscripted letters "i, j, k" are used to mark coreferential meaning; the dashed line "-->--" indicates that the previous sign is held in space with one hand when the other hand continues signing. For classifier predicates in sign languages, the gloss begins with the verb root printed in small letters, to be followed by the classifier handshape that the referent stands for. For example, 'be\_located+CL<sub>SASS</sub>' means that the verb root is a locative and the classifier is represented by a size-and-shape specifier handshape. Additionally, the speech is transcribed by using the romanization system of the target language, such as *Jyutping* for Cantonese.

(3b)

|        |                     |     |      |
|--------|---------------------|-----|------|
| *HKSL: | FATHER              | BOY | LIKE |
|        | "Daddy likes boys." |     |      |

(3c)

|       |                                            |                     |                        |                                |
|-------|--------------------------------------------|---------------------|------------------------|--------------------------------|
| HKSL: | LAST^NIGHT                                 | FATHER <sub>a</sub> | POLICEMAN <sub>b</sub> | <sub>a</sub> HELP <sub>b</sub> |
|       | "Last night, father helped the policeman." |                     |                        |                                |

(3d)

|       |                                            |        |                   |                    |
|-------|--------------------------------------------|--------|-------------------|--------------------|
| HKSL: | LAST^NIGHT                                 | FATHER | BOOK <sub>i</sub> | PUT <sub>a-i</sub> |
|       | "Last night, father put the book (there)." |        |                   |                    |

Note that agreement markings may be optional in HKSL; therefore, the order becomes SVO rather than SOV with uninflected agreement verbs (4a,b).

(4a)

|       |                               |      |     |                 |
|-------|-------------------------------|------|-----|-----------------|
| HKSL: | FATHER                        | HELP | BOY | IX <sub>a</sub> |
|       | "Daddy helps the boys there." |      |     |                 |

(4b)

|        |                               |     |                 |      |
|--------|-------------------------------|-----|-----------------|------|
| *HKSL: | FATHER                        | BOY | IX <sub>a</sub> | HELP |
|        | "Daddy helps the boys there." |     |                 |      |

### Analysis of Classifier Constructions in HKSL

There has been much debate about the grammatical status of classifier constructions. The iconic and mimetic nature of object and event depiction in classifier constructions has resulted in claims by some researchers that the term "classifier" is a misnomer. Instead, alternative terminologies have been suggested, such as "visual schematic representations" (Cogill-Koez, 2000), "depicting verbs/constructions" (Cormier et al., 2012), or "polycomponential verbs" (Schembri, 2003). Nonetheless, there are attempts to adopt a morphosyntactic analysis of classifier predicates in different sign languages. Supalla (1982, 1986) analyzing ASL proposes that classifier predicates are composed of movement roots and a set of affixes, among which handshapes and locations are obligatorily affixed to the verb stem and function as agreement markers. Within the framework of Minimalism, Benedicto and Brentari (2004) argue for the role of classifiers as morphosyntactic markers for external and internal arguments in transitive-intransitive and unergative-unaccusative alternations. They also posit that classifiers are heads of functional projections, i.e.,  $f_1P$  or  $f_2P$ , with morphosyntactic features which agree with those of an argument in the specifier position (i.e., structural agreement). Therefore, movement of an argument selected by the VP is either to an external argument position (i.e., Spec,  $f_1P$ ) or an internal argument position (i.e., Spec,  $f_2P$ ). However, unresolved issues remain, such as how body part classifiers and instrumental classifiers fit into the picture.

An alternative agreement analysis based on Distributed Morphology for classifier predicates is put forward by Glück

and Pfau (1998, 1999), who argue that both agreement verbs and classifying verbs share a similar morphological paradigm of agreement, in terms of moving between R-loci to show subject/object-verb agreement. But for classifying verbs there is another type of agreement, which is agreement between handshapes and the arguments they are denoting. This similarity is taken up in Zwitterlood (2003, 2008) who argues that classifiers have features for handshape and locus to spell out agreement in the structure<sup>3</sup>. At Numeration, the associated morphosyntactic feature bundles as well as a verb root are selected from List 1 and merged to form "root phrases" (rootPs) until a categorical "little  $\nu P$ ," a cyclic domain boundary for Spell-Out, is formed. This structure is shipped off to LF (Logic Form) for semantic interpretation and to PF for Vocabulary Insertion. At this stage, morphological operations apply on the PF branch, which is merger of agreement nodes for classifiers and R-loci, altering the syntactic structure hence word order changes accordingly. Vocabulary Items (i.e., elements from List 2) then compete for phonological realizations of the terminal nodes emerging from the syntactic structure. On the LF branch, the conceptual/intentional interface looks for interpretations for each terminal node (i.e., elements of List 3).

In this study, we will adopt the agreement analysis to account for the classifier constructions in HKSL and the related acquisition phenomenon. Following Zwitterlood (2003, 2008) and Glück and Pfau (1998, 1999), we assume there is agreement based on the handshape features and the antecedents; and at the same time, subject and object agreement can be spelt out through movement of the handshape classifiers between loci in space. At the descriptive level, classifier constructions in HKSL generally follow the schema of introducing the Ground before the Figure, as shown in (5a–5c)<sup>4</sup>. In other words, the Ground, like the theme NP in (5a) (i.e., BACKPACK), locative NP in (5b) (i.e., TOILET^ROLL), and goal NP in (5c) (i.e., TOY^CAR), is introduced into the discourse first through a locative predicate, with a classifier on the non-dominant hand being assigned to an R-locus in space. This classifier is sustained in space when the dominant hand introduces the Figure and a second classifier predicate, a phenomenon referred to as "perseveration." Note that in accounting for spatial expressions in NGT, Pfau and Aboh (2012) claim that the Part of the Ground (e.g., top of/next to the house) as expressed by H2 is usually left unexpressed. However, according to the native Deaf signers of HKSL, it is usually overt, consistently displaying a two-handed simultaneous, hence a Figure-Ground construction for the transitive predicate (5a), the locative existential predicate (5b), and the motion directional predicate (5c). Sometimes, the introduction of the object is simply by a topic (6a), or the object following the subject in an SOV order (6b).

<sup>3</sup>While person and number features have been confirmed in many studies, there have been discussions about what formal morphosyntactic features are there with the verbal classifiers in sign languages. Zwitterlood (2003) offers a detailed discussion on her proposal of classifiers bearing gender features.

<sup>4</sup>See Talmy (2000) for the Figure and Ground relations in spoken languages.

(5a)

|       |      |                                                                    |                                                          |                  |                                                                |
|-------|------|--------------------------------------------------------------------|----------------------------------------------------------|------------------|----------------------------------------------------------------|
| HKSL: | DH:  | BACKPACK <sub>i</sub>                                              |                                                          | CAT <sub>j</sub> | b <sub>push</sub> <sub>a</sub> +CL <sub>SEM</sub> <sub>j</sub> |
|       | NDH: |                                                                    | be_located <sub>a</sub> +CL <sub>SASS</sub> <sub>i</sub> | ----->-----      |                                                                |
|       |      | "The backpack is located here; the cat pushes it (with its side)." |                                                          |                  |                                                                |

(5b)

|       |      |                                                                             |                                                          |                       |                                                             |
|-------|------|-----------------------------------------------------------------------------|----------------------------------------------------------|-----------------------|-------------------------------------------------------------|
| HKSL: | DH:  | TOILET^ROLL <sub>i</sub>                                                    |                                                          | SCISSORS <sub>j</sub> | be_located_on <sub>a</sub> +CL <sub>SASS</sub> <sub>j</sub> |
|       | NDH: |                                                                             | be_located <sub>a</sub> +CL <sub>SASS</sub> <sub>i</sub> | ----->-----           |                                                             |
|       |      | "The toilet roll is located here; the pair of scissors is located (on) it." |                                                          |                       |                                                             |

(5c)

|       |      |                                                       |                                                         |                  |                                                                     |
|-------|------|-------------------------------------------------------|---------------------------------------------------------|------------------|---------------------------------------------------------------------|
| HKSL: | DH:  | TOY^CAR <sub>i</sub>                                  |                                                         | DOG <sub>j</sub> | b <sub>jump_onto</sub> <sub>a</sub> +CL <sub>SEM</sub> <sub>j</sub> |
|       | NDH: |                                                       | be_located <sub>a</sub> +CL <sub>SEM</sub> <sub>i</sub> | ----->-----      |                                                                     |
|       |      | "The toy car is located here; the dog jumps onto it." |                                                         |                  |                                                                     |

(6a)

|       |      |                               |                   |                     |                                                                 |
|-------|------|-------------------------------|-------------------|---------------------|-----------------------------------------------------------------|
|       |      | top                           |                   |                     |                                                                 |
| HKSL: | DH:  | IX <sub>a</sub>               | DOOR <sub>i</sub> |                     | a <sub>open</sub> <sub>b</sub> +CL <sub>HAND</sub> <sub>j</sub> |
|       | NDH: |                               |                   | MOTHER <sub>j</sub> | be_located <sub>a</sub> +CL <sub>SASS</sub> <sub>i</sub>        |
|       |      | "That door, mother opens it." |                   |                     |                                                                 |

(6b)

|       |      |                                        |                   |                                                                     |
|-------|------|----------------------------------------|-------------------|---------------------------------------------------------------------|
| HKSL: | DH:  | BOY <sub>i</sub>                       | VASE <sub>j</sub> | b <sub>kick</sub> <sub>a</sub> +CL <sub>BODYPART</sub> <sub>i</sub> |
|       | NDH: |                                        |                   | be_located <sub>a</sub> +CL <sub>SASS</sub> <sub>j</sub>            |
|       |      | "The boy kicks the vase with his leg." |                   |                                                                     |

Following Distributed Morphology and Zwitserlood (2003), we assume the root of a classifier predicate merges with different arguments bearing bundles of features to form rootPs, and eventually reaches a category node little *vP*, at which point the structure is shipped off for Spell-Out. At PF, the movement specification for the verb is inserted at the terminal node and different agreement projections are further merged above little *vP*. Subsequently, the feature bundles at Agr nodes, including the respective handshape and locus features for the Figure and Ground, are spelt out as classifier and spatial agreement markers via subject and object agreement respectively. Note that the arguments that are merged with the verb root vary in accordance with the predicate types. For (5a), the locative existential predicate "be\_located" requires a Theme and a Location argument and projects an AgrS and AgrIO nodes above little *vP*. For (5b), the motion directional predicate "jump" requires arguments for Theme, Source, and Goal and projects an AgrS and two AgrOO (oblique object) nodes. Finally, the transitive predicate "push" in (5c) requires an Agent argument for AgrS and a Theme argument for AgrDO. Basically, all the AgrS nodes will be spelt out and inserted with the phonological specification for classifiers of the Figure. This includes the external argument of unergative and transitives as well as the internal argument of unaccusative predicates at the specifier of AgrS. For (5a–c), the classifier presenting the Ground argument,

which refers to the object in an OSV order, is localized at an R-locus with which the movement of the Figure argument has to "agree" both in terms of spatial and grammatical agreement<sup>5</sup>. The phenomenon of perseveration shows that the classifier on H2 is an anaphoric expression which co-refers to the Ground argument introduced initially into the signing discourse. Although in the discourse the locative predicate following the Ground is omitted sometimes, the perseveration of the classifier on H2 at an R-locus is still observed in the predicate, like (6a). Therefore, we assume that the classifier on H2 inside the simultaneously articulated predicate is merged at the Spec position of the object agreement nodes and the syntactic derivation follows. The Figure may undergo movement to a functional projection higher than the Ground, to form the less frequently used SOV order like (6b).

To sum up, SVO order is not allowed in classifier constructions containing two noun referents in HKSL, while OSV based on a Ground-Figure schema is more frequently used than SOV. The descriptions above offer a framework to elucidate the syntactic function of moved or *in-situ* subjects and objects, as well as the status of classifiers as functional elements whose morphosyntactic features agree with the noun referents.

<sup>5</sup>Localization may involve a pointing sign, role shift, a locative existential predicate, or simply directing a classifier to an R-locus in space without a downward movement. Further research is necessary to figure out their syntactic consequences.

## Cantonese Counterparts of Classifier Constructions

Cantonese, though a classifier language, differs from HKSL in having numeral classifiers in the nominal as well as verbal domains. In (7a), *go3* is a nominal classifier and *kyun4* a verbal classifier. Also, verbs in Cantonese lack overt morphological agreement marking and grammatical relations are expressed primarily through the SVO order, as shown in the transitive (7a,b), locative existential (7c,d) and motion directional predicates (7e).

(7a)

|            |                                                             |           |      |      |            |            |       |              |
|------------|-------------------------------------------------------------|-----------|------|------|------------|------------|-------|--------------|
| Cantonese: | <b>go3</b>                                                  | naam4zai2 | daa2 | zo2  | <b>go3</b> | tung4hok4  | saam1 | <b>kyun4</b> |
|            | CL                                                          | boy       | hit  | PERF | CL         | schoolmate | three | CL           |
|            | "The boy punched the schoolmate three times with his fist." |           |      |      |            |            |       |              |

(7b)

|            |                                      |       |       |      |            |            |
|------------|--------------------------------------|-------|-------|------|------------|------------|
| Cantonese: | <b>zek3</b>                          | maau1 | teoi1 | gan2 | <b>go3</b> | syu1baau1  |
|            | CL                                   | cat   | push  | PROG | CL         | school-bag |
|            | "The cat is pushing the school bag." |       |       |      |            |            |

(7c)

|            |                                                        |      |           |             |      |             |                   |
|------------|--------------------------------------------------------|------|-----------|-------------|------|-------------|-------------------|
| Cantonese: | <b>jau5</b>                                            | baa2 | gaau3zin2 | <b>hai2</b> | gau6 | ci3zi2      | <b>soeng6min6</b> |
|            | have                                                   | CL   | scissors  | be located  | CL   | toilet roll | (on the) top of   |
|            | "A pair of scissors is on the top of the toilet roll." |      |           |             |      |             |                   |

(7d)

|            |                                                        |             |                   |             |      |           |
|------------|--------------------------------------------------------|-------------|-------------------|-------------|------|-----------|
| Cantonese: | gyun2                                                  | ci3zi2      | <b>soeng6min6</b> | <b>jau5</b> | baa2 | gaau3zin2 |
|            | CL                                                     | toilet roll | (on the) top of   | have        | CL   | scissors  |
|            | "A pair of scissors is on the top of the toilet roll." |             |                   |             |      |           |

(7e)

|            |                                              |      |                  |      |              |                   |
|------------|----------------------------------------------|------|------------------|------|--------------|-------------------|
| Cantonese: | zek3                                         | gau2 | tiu3 soeng6heoi3 | gaa3 | wun6geoi6ce1 | <b>soeng6min6</b> |
|            | CL                                           | dog  | jump up.go       | CL   | toy car      | (on the) top of   |
|            | "The dog jumps onto the top of the toy car." |      |                  |      |              |                   |

There are two alternative constructions for locative existentials in Cantonese. While maintaining an SVO order, (7c) uses a locative verb *hai2* "be located" and (7d) an existential verb *jau5* "have." Additionally, the locative NP is marked by a localizer<sup>6</sup> *soeng6min6* (on top of). Note that in the literature, a clause initial *jau5* "have" is analyzed as an existential quantifier introducing an indefinite NP *baa2 gaau3zin2* "a pair of scissors" into the discourse, as in (7c). In (7d), *jau5* "have" is analyzed as an existential verb selecting a locative NP as the grammatical subject (Huang, 1990). As for motion directional predicates, an SVO order maintains but the verbal domain is composed of serial verbs *tiu3 soeng6heoi3* "jump onto" in (7e).

<sup>6</sup>In Chinese, when verbs or prepositions select locations as their complements, it is necessary for the complement to take a localizer which denotes an axial part (Huang, 2009).

## Crosslinguistic Comparison and Acquisition Predictions

The grammatical descriptions above show crosslinguistic differences between HKSL and Cantonese regarding the three types of constructions (i.e., transitive, locative existential, and motion directional constructions), both in terms of word order and morphological complexity of the verb root. Classifier constructions in HKSL are primarily OSV, and sometimes SOV, while the equivalent constructions in Cantonese are

consistently SVO. Second, following Distributed Morphology, for the selection of morphosyntactic features from List 1, HKSL differs from Cantonese in the selection of roots, classifier features and locus features to mark subject/object as well as spatial agreement at the R-loci of the classifiers. As said, the selection of locus feature in the Numeration is crucial for spatial agreement function as they spell out the R-loci for the classifiers in space. Such properties are absent in Cantonese. Furthermore, locative existentials in Cantonese are explicitly encoded by a locative verb *hai2* "be located" or an existential verb *jau5* "have" and a localizer like *up*, whereas in HKSL such constructions require an abstract verb root *be\_located* and some placement affixes such as "next to" and "on top of" to encode the axial parts of the Ground entity with which the Figure sets up a spatial relation with. These crosslinguistic differences between HKSL and Cantonese pose

interesting acquisition predictions especially in the context of Deaf children acquiring HKSL in a bilingual fashion.

As discussed previously, the basic word order of HKSL is SVO with plain verbs, uninflected spatial and agreement verbs. Child data from Lam (2009) also confirmed an initial SVO order based on plain verbs. As such, it overlaps with the canonical SVO order in Cantonese. Under these circumstances, we predict that the initial word order of constructions involving a classifier predicate in HKSL is SVO, which may actually be doubly enhanced by the “shared” canonical SVO order of Cantonese and HKSL. Language Synthesis will predict that these DHs may initially select those morphosyntactic features pertaining to a SVO order with a lexicalized verb root, but not classifier features or locus features. Under those circumstances, it pertains to a Cantonese or a HKSL-based structure and the latter reflects the word order grammar of plain verbs and sometimes uninflected agreement verbs. As such, Vocabulary Insertion may come from Cantonese and HKSL, or both under code blending conditions.

Subsequent acquisition of inflectional morphology for person and spatial agreement with agreement verbs and spatial verbs may trigger Deaf children’s reanalysis of verb morphology, in the sense that HKSL verbs are not totally uninflected, leading to a reformulation of sub-classes of verbs and one of them is classifier constructions constituted by an abstract verb root, classifier features as well as locus features for spatial and subject/object agreement. We predict that classifier features are selected earlier than locus features in the Numeration, because classifier features, said to be akin to gender features in Zwitserlood (2003), are more semantic in nature, unlike locus features which yield R-loci in space for certain formal functions of encoding referential and agreement relations. The selection of such features in the Numeration motivates projections of agreement nodes at Spell-out where the features are merged at the terminal nodes for Spec-Head agreement with the noun referents in the specifier positions, and for spelling out the R-loci of the classifiers for subject/object agreement. In other words, the acquisition of the morphosyntactic properties of classifier constructions, and the schema of the Ground preceding the Figure in classifier constructions trigger Deaf children to develop word order variation, from SVO to OSV or SOV orders.

To sum up this section, we examine whether the selection of morphosyntactic features in the Numeration is a potential domain for language interaction to occur in our DHs’ production of HKSL classifier constructions. Lack of inhibition also implies that Vocabulary Insertion as a late phenomenon allows items to come from either Cantonese or HKSL.

## METHODOLOGY

### Participants

The current study involved 15 HKSL-Cantonese DHs who have been mainstreamed into a sign bilingual and co-enrollment (SLCO) environment in Hong Kong since kindergarten. The SLCO classes, comprised of Deaf and hearing students in a ratio of 1:3 or 1:4, are co-taught by a hearing teacher and a Deaf teacher who is either a native or a near-native signer of

HKSL. Totally, there are about 7 to 8 Deaf teachers in school who use primarily HKSL as the language of instruction and communication with other teachers and students, Deaf and hearing. The hearing teachers use primarily Cantonese and English, and sometimes Mandarin Chinese as the language of instruction; however, they also sign to facilitate communication whenever necessary. As both Deaf and hearing children are bimodal bilingual, they usually switch between Cantonese and HKSL in their daily interactions. At the time of the experiment, the DHs came from Primary 3 to Primary 6. Being DHs, the school is the only learning environment in which they receive consistent input in HKSL, in addition to Cantonese at home and at school. Note that they had HKSL exposure 1 h per week for 8–12 months before joining the SLCO Programme. In this study, we took the age of acquisition (AoA) of HKSL at the point when they started to receive consistent and ample input in HKSL in the SLCO Programme. At the time of the experiment, their chronological ages ranged from 8;10 to 14;5. Their AoA of HKSL ranged from 4;2 to 7;2. For five of these students, they could also be considered as late learners of HKSL due to exposure to the language at roughly age 6 or 7<sup>7</sup>.

We divided these 15 DHs into four groups on the basis of their duration of exposure to HKSL. Each group differed from the others by 1 year of exposure to HKSL. The DHs in Group 1 (aver. AoA of HKSL = 73.5 months) had the longest duration of exposure to HKSL for about 7 years. Those in Group 2 (aver. AoA of HKSL = 68.25 months), Group 3 (aver. AoA of HKSL = 55 months), and Group 4 (aver. AoA of HKSL = 59 months) had around 6, 5, and 4 years of exposure to HKSL, respectively. The numbers of DHs in each group were 4, 4, 4, and 3 for Groups 1, 2, 3, and 4, respectively. Among all of the 15 DHs, 11 of them have profound hearing loss (91+ dB), 3 of them are severely deaf (71–90 dB), and 1 have moderately severe hearing loss (56–70 dB). All of the 11 profoundly DHs are implanted, excluding 3 of them who wear hearing aids. Except for hearing loss, all of them do not have any other disabilities.

Two Deaf children of Deaf parents (DD-1 and DD-2), who are siblings to each other, took part in the current study as controls. DD-1 (studying with students of Group 1) is 1 year older than DD-2 (studying with children of Group 2). Due to misconception about sign language in HK earlier on, these two DDs did not have intensive HKSL exposure until 1;9 and 1;3 respectively; however, we suspect casual viewing of HKSL occurred at home since both of their parents are Deaf. DD-1 and DD-2 have been studying in the same SLCO Programme as the other 15 DHs. Their chronological ages were 12;9 and 11;3 respectively at the time of the experiment. **Table 1** summarizes the background information of the 15 DHs and 2 DDs.

For a better understanding of their knowledge of spoken languages, Cantonese and written Chinese assessments are

<sup>7</sup>A reviewer queried why the input before the SLCO Programme was not taken into consideration in this study. While these children joined a 45-minute sign language intervention programme weekly for at least one year before joining the SLCO Programme, they did not necessarily attend the sessions regularly because the Programme is not compulsory.

**TABLE 1** | Backgrounds of DHs and DDs.

| Participants | Research Code | Gender | Deaf parent(s) | Grade | Age (month) | HKSL AoA (month) | Duration of HKSL exposure (month) | Degree of hearing loss in the better ear (dB) | Hearing device | Age of wearing CI/HA (month) |
|--------------|---------------|--------|----------------|-------|-------------|------------------|-----------------------------------|-----------------------------------------------|----------------|------------------------------|
| Group 1      | DH-G1-1       | F      | No             | P6    | 140         | 52               | 88                                | 88                                            | CI             | 27                           |
|              | DH-G1-2       | F      | No             | P6    | 167         | 80               | 87                                | 118                                           | CI             | 41                           |
|              | DH-G1-3       | M      | No             | P6    | 163         | 76               | 87                                | 105                                           | CI             | 38                           |
|              | DH-G1-4       | M      | No             | P6    | 173         | 86               | 87                                | 108                                           | CI             | 71                           |
| Group 2      | DH-G2-1       | M      | No             | P5    | 129         | 54               | 75                                | 108                                           | CI             | 30                           |
|              | DH-G2-2       | M      | No             | P5    | 154         | 79               | 75                                | 107                                           | HA             | 38                           |
|              | DH-G2-3       | M      | No             | P5    | 150         | 66               | 84                                | 87                                            | HA             | 36                           |
|              | DH-G2-4       | F      | No             | P5    | 149         | 74               | 75                                | 120                                           | HA             | 3                            |
| Group 3      | DH-G3-1       | F      | No             | P4    | 122         | 54               | 68                                | 93                                            | HA             | 26                           |
|              | DH-G3-2       | F      | No             | P4    | 127         | 58               | 69                                | 97                                            | CI             | 23                           |
|              | DH-G3-3       | M      | No             | P4    | 127         | 58               | 69                                | 60                                            | HA             | 33                           |
|              | DH-G3-4       | F      | No             | P4    | 120         | 50               | 70                                | 108                                           | CI             | 24                           |
| Group 4      | DH-G4-1       | F      | No             | P3    | 119         | 63               | 56                                | 120                                           | CI             | 22                           |
|              | DH-G4-2       | F      | No             | P3    | 119         | 63               | 56                                | 120                                           | CI             | 22                           |
|              | DH-G4-3       | M      | No             | P3    | 106         | 51               | 55                                | 85                                            | HA             | 43                           |
| DD           | DD-1          | M      | Yes            | P6    | 153         | 21               | 132                               | 93                                            | HA             | 6                            |
|              | DD-2          | F      | Yes            | P5    | 135         | 15               | 120                               | 72                                            | HA             | 31                           |

**TABLE 2** | Deaf children's performance on spoken languages.

| Participants | Research Code | ACGK (%) | HKCOLAS-CG (%) | CLNT (%) | CanSWORT (%) |
|--------------|---------------|----------|----------------|----------|--------------|
| Group 1      | DH-G1-1       | 95.19    | 70.83          | 100      | 95           |
|              | DH-G1-2       | 84.62    | 48.38          | 8        | 4.17         |
|              | DH-G1-3       | 81.73    | 41.67          | 0        | 0            |
|              | DH-G1-4       | 81.73    | 32.87          | 68       | 78.33        |
| Group 2      | DH-G2-1       | 69.23    | 43.98          | 4        | 0            |
|              | DH-G2-2       | 83.65    | 52.78          | 72       | 55.83        |
|              | DH-G2-3       | 89.42    | 42.82          | 96       | 41.67        |
|              | DH-G2-4       | 91.35    | 33.10          | 0        | 0            |
| Group 3      | DH-G3-1       | 93.27    | 70.14          | 92       | 93.33        |
|              | DH-G3-2       | 79.81    | 31.71          | 84       | 85.83        |
|              | DH-G3-3       | 88.46    | 67.59          | 96       | 49.17        |
|              | DH-G3-4       | 95.19    | 56.71          | 100      | 95           |
| Group 4      | DH-G4-1       | 62.50    | 39.35          | 92       | 60.83        |
|              | DH-G4-2       | 75.96    | 43.06          | 84       | 93.33        |
|              | DH-G4-3       | 97.12    | 78.70          | 92       | 76.67        |
| DD           | DD-1          | 96.15    | 75.23          | 96       | 91.67        |
|              | DD-2          | 92.31    | 66.67          | 92       | 100          |

administered to the 15 DHs and 2 DDs annually, which are the *Assessment of Chinese Grammatical Knowledge (ACGK)*, and the subscale on *Cantonese Grammar of Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS-CG)* (T'sou et al.,

2006). ACGK is an unpublished assessment tool developed by the Centre for Sign Linguistics and Deaf Studies, Chinese University of Hong Kong. It aims to assess children's syntactic and morpho-syntactic knowledge of written Chinese that is based on Mandarin Chinese grammar. HKCOLAS-CG is a standardized tool for assessing children's grammatical knowledge of spoken Cantonese. All test items in ACGK are presented in written Chinese whereas HKCOLAS-CG requires children to listen and make responses in Cantonese. Since the Deaf children's speech perception abilities varied, the low scores that some achieved in HKCOLAS-CG may be due to the auditory mode of the assessment. **Table 2** lists each participant's scores of ACGK and HKCOLAS-CG, which were obtained during the same time when the current study was conducted. Their speech perception scores were collected based on two Cantonese assessment tools, one for tone identification—*Cantonese Lexical Neighborhood Test (CLNT)* (Yuen et al., 2008) and the other one for disyllabic word recognition—*Cantonese Spoken Word Recognition Test (CanSWORT)* (Ng, 2014). Note also that Tang et al. (2014) reported a significant positive correlation not only between 20 SLCO Deaf children's developing grammatical knowledge of oral Cantonese and written Chinese ( $r = 0.790^{**}$ ,  $p = 0.000$ , 1-tailed); but also a positive interaction between HKSL and written Chinese ( $r = 0.591^{**}$ ,  $p = 0.003$ , 1-tailed) and between HKSL and oral Cantonese ( $r = 0.663^{**}$ ,  $p = 0.001$ , 1-tailed). The data analyzed in Tang et al. (2014) came from the same assessment tools mentioned here, including ACGK, HKCLOS-C as well as *Hong Kong Sign Language Elicitation Tool (HKSL-ET)*. Meanwhile, all the DHs and DDs in the current study, except for 1 DH in Group 3 and 1 in Group 4, were subjects in Tang et al. (2014).

In this experiment, three native Deaf signers (1 male and 2 female) participated as controls. All of them had two signing Deaf parents. They were 27-, 28-, and 33-year-old at the time of the experiment. Two of them graduated from the same school for the deaf that adopted the oral approach. The third one attended the same deaf school as the other two but transferred to a mainstream secondary school from Form 4 to Form 7.

## Materials and Elicitation Procedures

This study was part of a large-scale project approved by the Survey and Behavioral Research Ethics Committee (SBREC) at The Chinese University of Hong Kong. All the adult participants and parents of child participants signed a written, informed consent form. The child participants were individually tested in a quiet room at school while the adult participants were tested at the Centre for Sign Linguistics and Deaf Studies. Trained Deaf research assistants followed a strict protocol when administering the test battery, *Hong Kong Sign Language Elicitation Tool*, which is an unpublished assessment tool for profiling Deaf children's HKSL development in terms of production and judgments of grammaticality. The tool includes several subtests for different grammatical components, including classifier constructions, agreement verbs, negators, modals, *wh*-questions, *yes-no* questions, and non-manual adverbials.

The test on classifier constructions was a picture description task which took about 15 min to complete. In this task, all participants were asked to describe a set of 16 pictures in HKSL: six pictures for locative existential constructions, six for motion directional constructions and four for transitive classifier constructions. **Figure 1** provides three sample pictures as stimuli for eliciting the different types of classifier constructions in the current study. The target HKSL sentences can be seen in examples (5a–c), while (7b–e) are the Cantonese counterparts. The experimenter showed the pictures one by one to the participants, who were allowed time to study the picture. Then, the experimenter removed the stimuli and the participants described the picture in HKSL. Additionally, a picture-naming task was conducted prior to the picture description task to control for vocabulary comprehension, as lexical variation is common among the HKSL signers, so a vocabulary check was necessary to ensure the participants' comprehension and production of the objects in the stimuli. The whole procedure was video-taped, and the participants' productions of the stimuli were transcribed using ELAN and coded accordingly.

As pointed out in the previous sections, a change of word order occurs with complex verb morphology in a classifier construction. In this study, all stimuli involved two arguments mapped onto a grammatical subject and object respectively. We selected different predicate roots, phonologically expressed by the dominant hand moving toward the non-dominant hand. In locative existential predicates, the locative root “be\_located” requires a small downward movement toward a location argument. In motion directional predicates, three transfer roots—“jump onto,” “fall from,” and “fall onto”—were selected for the experiment. They require an “arc” path movement of the dominant hand from one R-locus to another R-locus that is

occupied by the non-dominant hand. The transitive predicates also involve a transfer root translated as “push” and “press against.” It involves a path or orientation change of the dominant hand toward the non-dominant hand.

There are three types of classifiers in the predicates, coded based on Supalla (1982) categorization—semantic, SASS (i.e., size-and-shape specifiers), and bodypart classifiers. The semantic classifiers were used for co-reference with a dog, a cat, an elephant, a horse, and a toy car; SASSes for a rock, a backpack, a present, a toilet roll, and a pair of scissors; and bodypart classifier for a bionic hand. These classifiers were assigned to the dominant hand in the formation of a locative existential or motion directional predicate, where the non-dominant hand was either a semantic classifier or a SASS classifier. For the transitive predicates, only semantic and SASS classifiers were adopted. The classifiers on the dominant hand were all semantic, while the classifiers on the non-dominant hand were either SASS or semantic classifiers.

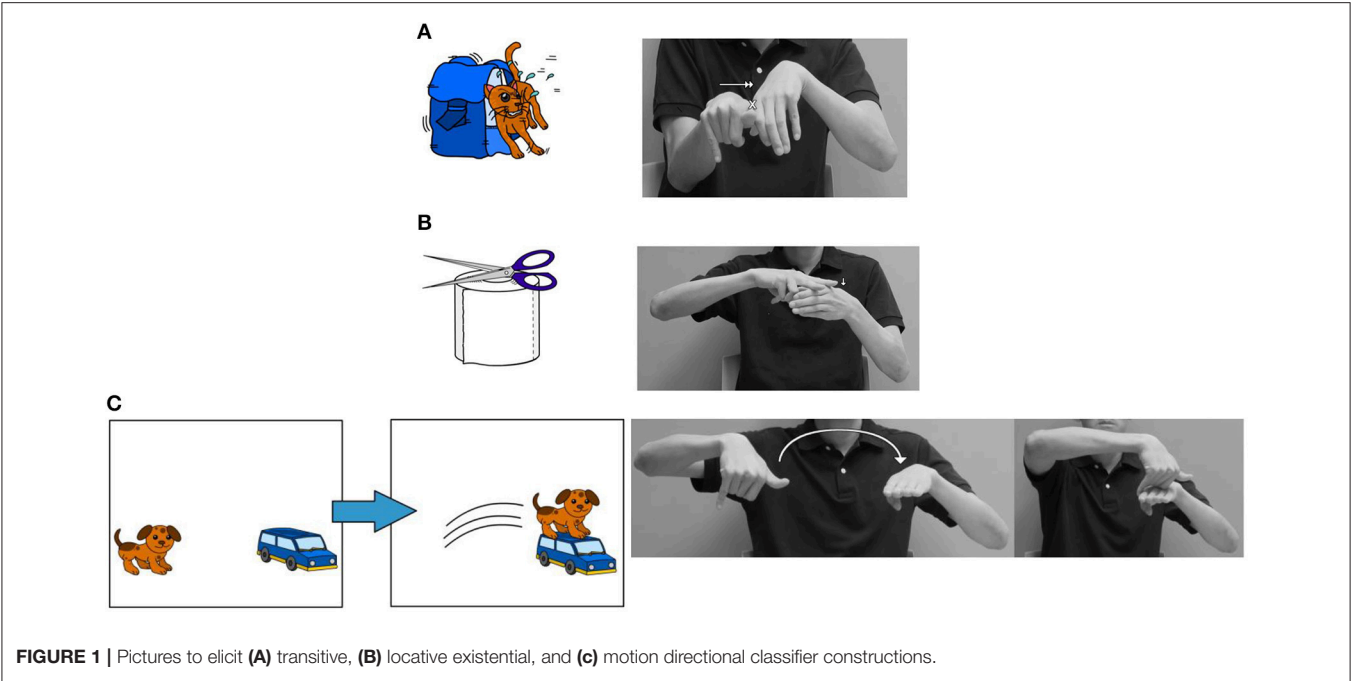
## Coding Procedures

All production data were transcribed using ELAN (<http://tla.mpi.nl/tools/tla-tools/elan/>; Crasborn and Sloetjes, 2008) and coded with reference to a set of criteria based on reported analyses of HKSL. In this paper, two criteria were adopted in coding the children's performance. The first one was *verb root* of the main predicate, realized phonologically by the movement of the classifier on the dominant hand toward that on the non-dominant hand (henceforth MVR). The second one was *word order* (henceforth WO). We focused on these two criteria because we predict that properties of the verb root interact with word order changes in classifier constructions. Using the adults' performance as controls, the Deaf children's productions were categorized into adult-like performance and non-adult-like performance. The children's encoding of the predicates through gesture, lexical verbs, classifying verbs comprised of classifier handshapes was also coded. The data were scored by one Deaf researcher who is a native signer of HKSL, and one hearing researcher who is one of the co-authors of this paper. The rate of agreement between the two coders on the two criteria was 90%.

## RESULTS

### Adult Deaf Signers

Data from three adult native Deaf signers formed the baseline of the current study. All of their responses showed adult-like classifier constructions in terms of target MVR and WO, except for one token of WO (see 8). Instead of one motion directional classifier construction, a male Deaf participant produced a serial verb construction made up of a locative existential classifier predicate and two motion directional classifier predicates. Such kind of serial verb constructions was seldom observed in the Deaf children's data. In all, data from the adult Deaf signers suggested the stimuli for the current study are sensitive to eliciting classifier constructions.



(8)

**FIGURE 1 |** Pictures to elicit (A) transitive, (B) locative existential, and (C) motion directional classifier constructions.

|       |      |                                                                                                                                                      |                                             |                       |                                                   |
|-------|------|------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|-----------------------|---------------------------------------------------|
| HKSL: | DH:  | DOG <sub>i</sub>                                                                                                                                     |                                             | SCISSORS <sub>j</sub> | be_located_on <sub>b</sub> +CL <sub>SASSj</sub>   |
|       | NDH: |                                                                                                                                                      | be_located <sub>a</sub> +CL <sub>SEMI</sub> | ----->-----           |                                                   |
|       | DH:  |                                                                                                                                                      | ----->-----                                 |                       | b_fall_off_from <sub>d</sub> +CL <sub>SASSj</sub> |
|       | NDH: |                                                                                                                                                      | b_move <sub>c</sub> +CL <sub>SEMI</sub>     | ----->-----           |                                                   |
|       |      | "A dog is located here; a pair of scissors is on (the back of) the dog; the dog moves and the pair of scissors falls off from (the back of) the dog" |                                             |                       |                                                   |

As mentioned, while classifier constructions allow both OSV and SOV orders, the former order is much more common than the latter. This is confirmed by our adult signers' productions (Table 3). Over 94% of the tokens produced were of an OSV order (Table 3). Only 1 token of SOV order with a locative existential predicate was found.

### Deaf Children

Using the results from the adult signers as baseline, we coded the responses as non-adult-like performance when no classifier construction was produced by the DHs. One DH from Group 1 actually produced three tokens of transitive predicates in an SVO order with no classifier constructions (see 9); however, these sentences were coded as grammatical. According to the

Deaf rater, the child used role shift together with an inflected agreement verb PUSH in the main predicate. Therefore, we removed these 3 tokens from our analysis. Table 4 summarizes the distribution of the total number of responses (Group x Types of classifier constructions x Number of tokens). Subsequent analyses presented below adopt these numbers as denominators in the calculation.

(9)

|       |                             |                    |                   |                         |
|-------|-----------------------------|--------------------|-------------------|-------------------------|
|       |                             | _____rs            |                   |                         |
| HKSL: | CAT <sub>a</sub>            | aPUSH <sub>b</sub> | IX <sub>b-j</sub> | ELEPHANT <sub>b-j</sub> |
|       | "A cat pushes an elephant." |                    |                   | (DH-G1-3)               |

Generally, while all tokens of WO from the DDs were adult-like (i.e., 100%), only 66% of their MVR tokens were adult-like (see Table 5). As for the DHs, the numbers of adult-like tokens of MVR and WO of Group 1 were similar to those DDs (i.e., MVR = 74%; WO = 92%), suggesting the possibility of achieving near-native competence in classifier constructions. On the other hand, the number of adult-like responses for MVR and WO dropped from Groups 2 to 4. Group 2's adult-like responses were 56% for MVR and 58% for WO; Group 3 were 39% for MVR and

**TABLE 3 |** Performance on WO by adult Deaf signers.

| Predicates | O > S > V (%) | S > O > V (%) | Serial verbs (%) | Total no. of responses (N = 3) |
|------------|---------------|---------------|------------------|--------------------------------|
| Transitive | 12 (1.00)     | –             | –                | 12                             |
| Loc-exist  | 17 (.94)      | 1(0.06)       | –                | 18                             |
| Mot-dir    | 17 (.94)      | –             | 1 (0.06)         | 18                             |

**TABLE 4 |** Number of responses for the current analyses\*.

| Participants            | Transitive | Loc-exist | Mot-dir | *Total number of responses (%) |
|-------------------------|------------|-----------|---------|--------------------------------|
| Group 1 ( <i>N</i> = 4) | 13         | 24        | 24      | 61/64 (0.95)                   |
| Group 2 ( <i>N</i> = 4) | 16         | 24        | 24      | 64/64 (1.00)                   |
| Group 3 ( <i>N</i> = 4) | 16         | 24        | 24      | 64/64 (1.00)                   |
| Group 4 ( <i>N</i> = 3) | 12         | 18        | 18      | 48/48 (1.00)                   |
| DD ( <i>N</i> = 2)      | 8          | 12        | 12      | 32/32 (1.00)                   |

22% for WO; and Group 4 were 46% for MVR and 21% for WO. These results suggest that duration of exposure to HKSL has an effect on their acquisition. Also, Groups 3 and 4's performance on WO implies that word order changes in classifier constructions posed initial difficulty. In the following two sections, we will describe the participants' performance on MVR and WO.

## Deaf Children's Performance on MVR

As mentioned, the verb root of a classifier predicate is morphologically different from the other types of verbs in sign languages. **Table 6** shows the distribution of adult-like MVR responses over the three types of classifier constructions. While almost all DHs reached the ceiling of performance on MVR in motion directional predicates, their production of adult-like MVR in transitive and locative existential predicates dropped dramatically. The MVR of locative existential predicates turned out to be the most difficult for all children, including the DDs<sup>8</sup>. As shown in **Table 6**, a great majority of them, especially those in Groups 3 and 4, either failed to produce a classifier predicate and used other lexical verbs (e.g., HAVE) or failed to realize the verb "be\_located" using a small downward movement. In the latter case, they adopted a long downward movement which bears other predicate meanings (also see data description below). Previous studies argued that due to iconicity, not only DDs but also DHs can spatially encode the locative relation between a Figure and a Ground as early as age 2;0 (Lindert, 2001). The current findings suggest locating them at specific R-loci in space through a specific movement feature turned out to be quite difficult. We argue that it is due to their not selecting the locus features from List 1 initially, and at the same time not realizing that the properties of movement are morphemic.

To further analyze group performance on the verb root, we categorized the DHs' errors into two types (see **Figures 2A,B** and **Table 7**). The first type of errors shows the DHs' lack of production of classifier predicates (i.e., "No CL-pred"). As shown, such a lack was observed only in transitive and locative existential predicates but not motion directional predicates, especially among children from Groups 3 and 4. These children selected an equivalent lexical verb instead if they could identify

one, such as PUSH in (10a) and HAVE in (10b). Note that the agreement verb PUSH in the context of an ordinary transitive predicate requires an SOV/OSV order or in SVO order with role shift, as in (9). However, none of such word orders or role shift with SVO order was observed in the DH's productions.

(10a)

|        |                             |      |           |
|--------|-----------------------------|------|-----------|
| *HKSL: | CAT                         | PUSH | ELEPHANT  |
|        | "A cat pushes an elephant." |      | (DH-G3-4) |

(10b)

|        |                                             |                  |      |           |
|--------|---------------------------------------------|------------------|------|-----------|
| *HKSL: | TOILET^ROLL                                 | IX <sub>up</sub> | HAVE | SCISSORS  |
|        | "A pair of scissors is on the toilet roll." |                  |      | (DH-G4-2) |

In total, there were 50 tokens of MVR errors under the category of "No CL-pred" among which a majority of them (43 out of 50 tokens, 86%) showed a SVO order and involved either a lexical verb or gesture (see section SVO Order With a Variety of Verb Roots, **Table 10**). It is obvious that these children resorted to selecting a lexical verb root initially, and the lower the grades the higher the percentages of such erroneous productions. Therefore, so far as the transitive predicates and locative existential predicates are concerned, Deaf children from the lower grades tended to select, from List 1, a lexical root but not features pertaining to a classifier construction in the Numeration.

The second type of errors is related to how children encode events or states realized by movement (i.e., verb root) with classifier morphemes. In our analysis, we assumed such errors were morpho-phonological (i.e., "Non-target MVR" in **Figure 3** and **Table 7**), and were generally found in the locative existential predicates. In fact, all the errors produced by the DDs belonged into this category. For the native Deaf adults we consulted, these non-target MVRs encode a different predicate meaning. As said above, most DHs and DDs produced a long downward movement for locative existential predicates instead of the target which is a small downward movement. Such a long downward movement signals three different meanings: "fall down from (a high position)," "put something at (a location)," and "jump onto" a location. Among the 44 tokens of such errors extracted from the locative existential predicates, about 28 of them produced by the DHs had a meaning of "put something at (a location)," and 7 such tokens were accompanied by mouthing the Cantonese verbs *fong3* or *baai2* "put." This finding suggests that, instead of selecting an abstract HKSL verb root "be\_located," these children preferred to select a lexical, locative verb like *fong3* in Cantonese (e.g., *baa2 gaau3zin2 fong3 hai6 gyun2 ci3zi2 soeng6min6* "The scissors are (placed) on top of the toilet roll").

In sum, the findings of MVR reveal that Deaf children experienced initial difficulty in selecting an abstract verb root for the classifier predicates in HKSL. Before converging on the adults' grammar, we observed a lack of use of classifier predicates, especially in locative existential predicates, and insertion of a lexical verb root was the usual strategy, if they could identify one. Also, adopting an appropriate movement shape to encode the existential verb root led to morphological errors in their production. In the next section, we proceed to analyze how Deaf

<sup>8</sup>A reviewer asked if the consistent difficulty in encoding the verb root 'be\_located' lends support to Koulidobrova's (2016) observation of object omission. In our data, objects were usually not omitted despite the absence of 'be\_located.' The DHs actually adopted an alternative strategy to encode the verb root (e.g. a lexical verb HAVE) having a similar meaning instead (see **Table 7**)

**TABLE 5 |** Production of adult-like MVR/WO by DHs and DDs.

| Participants            | Predicates | Adult-like MVR | Group Total (%) | Adult-like WO | Group Total (%) | Total no. of responses |
|-------------------------|------------|----------------|-----------------|---------------|-----------------|------------------------|
| Group 1 ( <i>N</i> = 4) | Transitive | 10             | 45 (0.74)       | 12            | 56 (0.92)       | 61                     |
|                         | Loc-exist  | 11             |                 | 22            |                 |                        |
|                         | Mot-dir    | 24             |                 | 22            |                 |                        |
| Group 2 ( <i>N</i> = 4) | Transitive | 9              | 36 (0.56)       | 9             | 37 (0.58)       | 64                     |
|                         | Loc-exist  | 3              |                 | 12            |                 |                        |
|                         | Mot-dir    | 24             |                 | 16            |                 |                        |
| Group 3 ( <i>N</i> = 4) | Transitive | 0              | 25 (0.39)       | 4             | 14 (0.22)       | 64                     |
|                         | Loc-exist  | 1              |                 | 6             |                 |                        |
|                         | Mot-dir    | 24             |                 | 4             |                 |                        |
| Group 4 ( <i>N</i> = 3) | Transitive | 4              | 22 (0.46)       | 1             | 10 (0.21)       | 48                     |
|                         | Loc-exist  | 0              |                 | 1             |                 |                        |
|                         | Mot-dir    | 18             |                 | 8             |                 |                        |
| DD ( <i>N</i> = 2)      | Transitive | 5              | 21 (0.66)       | 8             | 32 (1.00)       | 32                     |
|                         | Loc-exist  | 4              |                 | 12            |                 |                        |
|                         | Mot-dir    | 12             |                 | 12            |                 |                        |

**TABLE 6 |** Production of adult-like MVR by DHs and DDs.

| Predicate types | Transitive               |                 | Loc-exist                |                 | Mot-dir                  |                 | Adult-like responses/total responses per group (%) |
|-----------------|--------------------------|-----------------|--------------------------|-----------------|--------------------------|-----------------|----------------------------------------------------|
|                 | Adult-like responses (%) | Total responses | Adult-like responses (%) | Total responses | Adult-like responses (%) | Total responses |                                                    |
| Group 1         | 10 (0.77)                | 13              | 11 (0.46)                | 24              | 24 (1.00)                | 24              | 45/61 (0.74)                                       |
| Group 2         | 9 (0.56)                 | 16              | 3 (0.13)                 | 24              | 24 (1.00)                | 24              | 36/64 (0.56)                                       |
| Group 3         | 0 (0.00)                 | 16              | 1 (0.04)                 | 24              | 24 (1.00)                | 24              | 25/64 (0.39)                                       |
| Group 4         | 4 (0.33)                 | 12              | 0 (0.00)                 | 18              | 16 (.89)                 | 18              | 20/48 (0.42)                                       |
| DD-1            | 1 (0.25)                 | 4               | 2 (0.33)                 | 6               | 6 (1.00)                 | 6               | 9/16 (0.56)                                        |
| DD-2            | 4 (1.00)                 | 4               | 2 (0.33)                 | 6               | 6 (1.00)                 | 6               | 12/16 (0.75)                                       |

children's knowledge of verb root interacts with their acquisition of word order.

## Deaf Children's Performance on Word Order

As said, while OSV and SOV are the two acceptable word orders of classifier constructions in HKSL, elicited data from three native adult signers showed that OSV order was more prevalent, except for 2 tokens (see **Table 3**). As for the Deaf children, **Table 8** shows that the two DDs produced adult-like word order consistently. Additionally, 117 out of 237 responses of the DHs were adult-like; and those DHs with longer exposure to HKSL produced more tokens of adult-like word order for all three types of classifier constructions. Group 1 reached almost the ceiling of performance (i.e., 92%), Group 2 between 50 and 67%, but Groups 3 and 4 had much fewer adult-like WOs for all three types of classifier constructions.

### OSV Order as the Preferred Word Order

**Figures 3A,B** as well as **Table 9** show the different word order produced by the DHs. Similar to the native adults, all of the

32 WO responses produced by the DDs reflected the adult-like OSV order (see **Figure 3B**). **Figure 3A** shows that there was a big tendency for the OSV order in the DH's adult-like productions. This preference was observed even among the DHs of Groups 3 and 4. In fact, there were very few tokens of SOV order in the data, suggesting that it was a much less preferred order among the DDs and DHs. Also, the production of an OSV order for classifier constructions by Deaf children, as we argue, is taken to be evidence that they are selecting the locus features for the classifiers in the Numeration, for them to assign the classifier for the grammatical object to an R-locus in space through an initial locative existential predicate.

### Non-adult-like Responses

#### *SVO order with a variety of verb roots*

Among the 120 non-adult-like productions out of the 237 responses produced by the DHs, 91 (i.e., 76%) reflected a clear SVO order which is not acceptable for classifier constructions. In fact, it is difficult to determine if the knowledge of SVO order stems from Cantonese or HKSL,

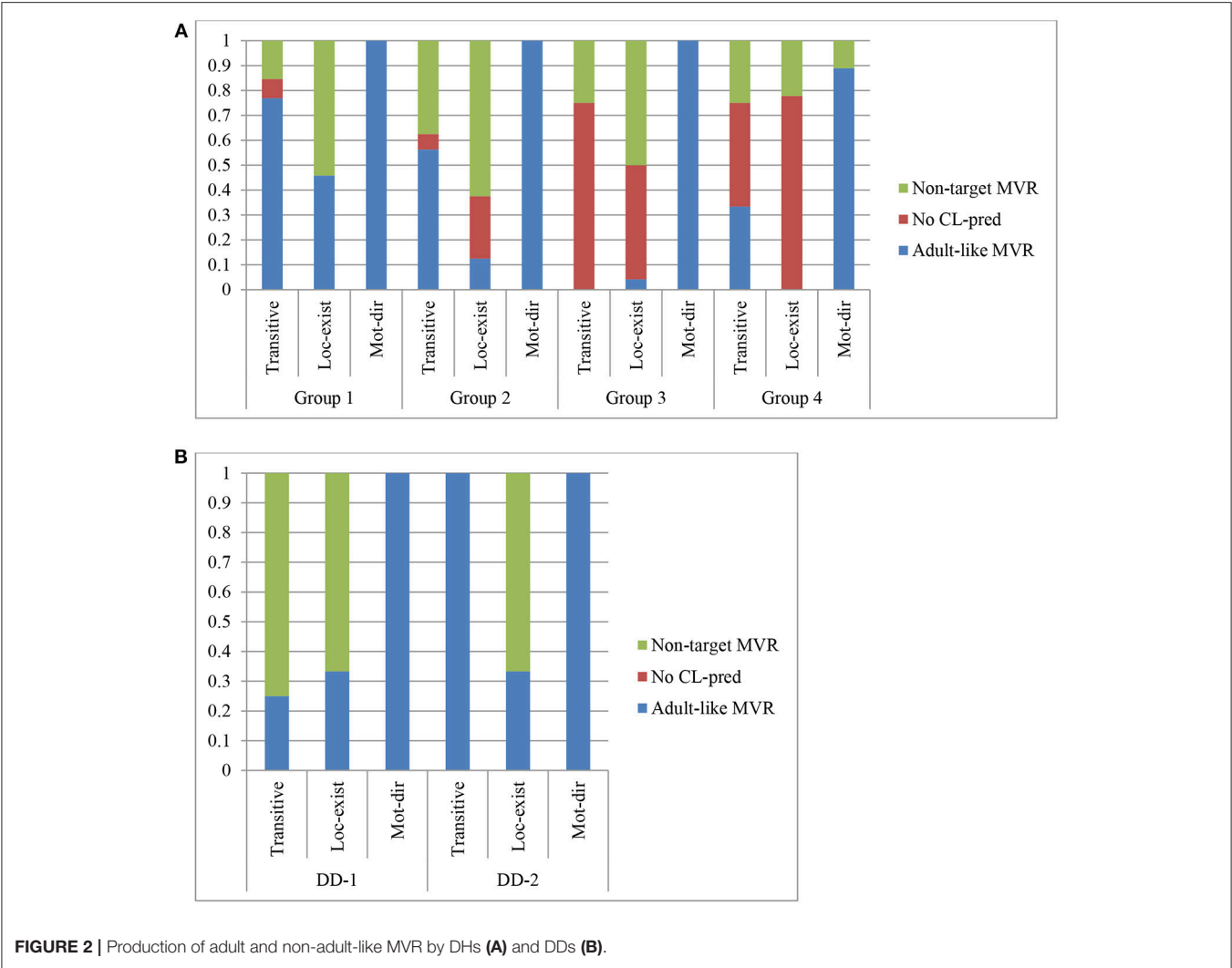


**FIGURE 2 |** Production of adult and non-adult-like MVR by DHs (A) and DDs (B).

**TABLE 7 |** Production of non-adult-like MVR by DHs and DDs.

| Predicate types | Transitive     |                    | Loc-exist      |                    | Mot-dir        |                    |
|-----------------|----------------|--------------------|----------------|--------------------|----------------|--------------------|
|                 | No CL-pred (%) | Non-target MVR (%) | No CL-pred (%) | Non-target MVR (%) | No CL-pred (%) | Non-target MVR (%) |
| Group 1         | 1 (0.08)       | 2 (0.15)           | 0 (0.00)       | 13 (0.54)          | 0 (0.00)       | 0 (0.00)           |
| Group 2         | 1 (0.06)       | 6 (0.38)           | 6 (0.25)       | 15 (0.63)          | 0 (0.00)       | 0 (0.00)           |
| Group 3         | 12 (0.75)      | 4 (0.25)           | 11 (0.46)      | 12 (0.50)          | 0 (0.00)       | 0 (0.00)           |
| Group 4         | 5 (0.42)       | 3 (0.25)           | 14 (0.78)      | 4 (0.22)           | 0 (0.00)       | 2 (0.11)           |
| DD-1            | 0 (0.00)       | 3 (0.75)           | 0 (0.00)       | 4 (0.67)           | 0 (0.00)       | 0 (0.00)           |
| DD-2            | 0 (0.00)       | 0 (0.00)           | 0 (0.00)       | 4 (0.67)           | 0 (0.00)       | 0 (0.00)           |

as both languages allow SVO as the basic word order, as discussed previously (see section Crosslinguistic Comparison and Acquisition Predictions). Yet, the way these children inserted the verb root into this basic SVO structure deserves our attention. We found 5 types of “verb roots” from their production (see Table 10). 68% of such errors belonged to either uninflected lexical verbs (i.e.,  $V_{\text{lexical}}$ ) or a form of two-handed signs which did not resemble a lexical sign. They were usually configured by two inappropriate classifier-like handshapes

(i.e.,  $V_{\text{complex}}$ , see Figure 4) and without spatial information. We took such productions to be morphologically complex signs but non-target both in terms of handshape configuration and spatial information. Other types of verb roots were just 4 tokens of gesture [see (11) produced by DH-G3-1], 2 tokens of verb series  $V_{\text{complex}} + V_{\text{lexical}}$ , and 2 tokens of a one-handed motion directional predicate. Following Language Synthesis, the  $S > V_{\text{lexical}} > O$  structure represents an output based on selecting the morphosyntactic features pertaining to a lexical

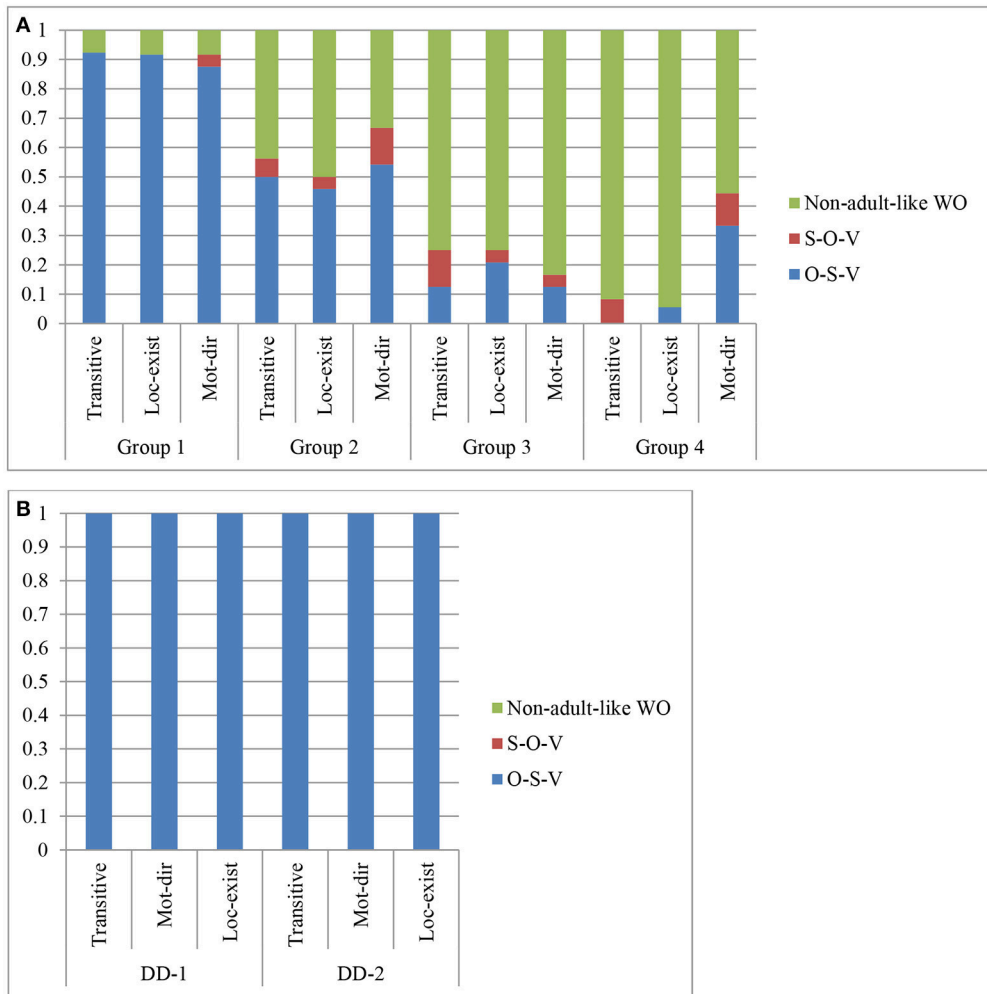


FIGURE 3 | Production of adult and non-adult-like WO by DHs (A) and DDs (B).

verb root without classifier or locus features leading to PUSH or PUT in Vocabulary Insertion. This phenomenon occurred more frequently with the DHs in Group 3 and 4 but gradually dropped upon longer duration of exposure to HKSL.

(11)

|        |                                                         |                               |       |           |
|--------|---------------------------------------------------------|-------------------------------|-------|-----------|
| *HKSL: | SMALL^CAT                                               | gesture [= push by shoulders] | HEAVY | ELEPHANT  |
|        | "A small cat pushes (itself against) a heavy elephant." |                               |       | (DH-G3-1) |

While  $V_{\text{lexical}}$  predominated the data of Groups 3 and 4,  $V_{\text{complex}}$  showed an almost reverse pattern of distribution, in the sense of an increasing tendency of production when the DHs moved up to Grades 2 and 1. In Group 1, the DHs knew SVO with a  $V_{\text{complex}}$  was ungrammatical in HKSL, as evidenced by the production of just two tokens of  $S > V_{\text{complex}} > O$ . In other words, the production of a  $V_{\text{complex}}$  sign during the initial acquisition process did not necessarily trigger reordering of SVO, contrary to our prediction. One reason is that when knowledge of SVO order based on a lexical verb root is doubly enhanced by Cantonese

and HKSL, these children might wrongly assume that verbs are paradigmatically lexical in nature. Another reason may stem from ambiguous input. We suspect that the DHs from Groups 2, and especially 3 and 4, might initially produce these  $V_{\text{complex}}$

signs as "lexical signs," similar to those two-handed lexical verb signs like SCOLD or REBEL in HKSL (Figure 5), which do not bear any locus or classifier features although they have a classifier predicate origin. Therefore, the erroneous constructions suggest that projections for object agreement which triggers word order changes were not in place yet, due to the absence of locus features despite the presence of classifier features. Consequentially, the word order remained as SVO as no formal agreement relation was established between the verb and the R-loci (see Discussion below).

**TABLE 8 |** Production of adult-like WO by DHs and DDs.

| Predicate types | Transitive               |                 | Loc-exist                |                 | Mot-dir                  |                 | Adult-like responses /total responses per group (%) |
|-----------------|--------------------------|-----------------|--------------------------|-----------------|--------------------------|-----------------|-----------------------------------------------------|
|                 | Adult-like responses (%) | Total responses | Adult-like responses (%) | Total responses | Adult-like responses (%) | Total responses |                                                     |
| Group 1         | 12 (0.92)                | 13              | 22 (0.92)                | 24              | 22 (0.92)                | 24              | 56/61 (0.92)                                        |
| Group 2         | 9 (0.56)                 | 16              | 12 (0.50)                | 24              | 16 (0.67)                | 24              | 37/64 (0.58)                                        |
| Group 3         | 4 (0.25)                 | 16              | 6 (0.25)                 | 24              | 4 (0.17)                 | 24              | 14/64 (0.22)                                        |
| Group 4         | 1 (0.08)                 | 12              | 1 (0.06)                 | 18              | 8 (0.44)                 | 18              | 10/48 (0.21)                                        |
| DD-1            | 4 (1.00)                 | 4               | 6 (1.00)                 | 6               | 6 (1.00)                 | 6               | 16/16 (1.00)                                        |
| DD-2            | 4 (1.00)                 | 4               | 6 (1.00)                 | 6               | 6 (1.00)                 | 6               | 16/16 (1.00)                                        |

**Table 11** offers a further analysis of the distribution of the two major non-adult-like verb roots,  $V_{\text{lexical}}$  and  $V_{\text{complex}}$ , in SVO order. The data are organized based on the DHs' performance on the types of classifier predicates by groups.

As for  $V_{\text{lexical}}$ , uninflected PUSH is selected consistently in transitive predicates. For locative existential predicates, we found a variety of lexical verbs such as uninflected PUT and main verb HAVE (see 12, 13a). These verbs were usually accompanied by a pointing sign  $IX_{\text{up}}$  that served more like a Cantonese localizer *soeng6min6* "up." The use of HAVE, as in (13a), has a meaning similar to the existential verb *yau5* "have" in Cantonese [see (13b) for the Cantonese counterpart], suggesting that the structure with PUT or HAVE is based on the Cantonese SVO order (see next section).

Note that in **Table 11**, we found no records of DHs across all groups inserting a  $V_{\text{lexical}}$  into a motion directional predicate in HKSL. The lack of equivalence in the morphosyntactic structure of verb roots between HKSL and Cantonese may be at play here. In HKSL, the verb root is expressed morphophonologically by a single path movement, which also iconically maps the path between the source and the goal arguments; however, Cantonese's motion directional predicates require serial verb constructions, such as *tiu3soeng6heoi3* (lit. "jump ascend go"), *dit3lok6lei4* (lit. "fall descend come"). It is interesting to observe that the DHs seemed to be sensitive to such differences early on, as evidenced by a high instance of correct MVR tokens (see **Figure 2A**).

(12)

|        |                                           |                       |     |        |                          |                  |
|--------|-------------------------------------------|-----------------------|-----|--------|--------------------------|------------------|
| *HKSL: | $IX_i$                                    | SCISSORS <sub>i</sub> | PUT | $IX_j$ | TOILET^ROLL <sub>j</sub> | $IX_{\text{up}}$ |
|        | "A pair of scissors is on a toilet roll." |                       |     |        |                          | (DH-G4-3)        |

(13a)

|        |                              |                  |      |             |
|--------|------------------------------|------------------|------|-------------|
| *HKSL: | DOG                          | $IX_{\text{up}}$ | HAVE | BIONIC^HAND |
|        | "A bionic hand is on a dog." |                  |      | (DH-G3-4)   |

(13b)

|            |                              |      |                 |      |      |              |
|------------|------------------------------|------|-----------------|------|------|--------------|
| Cantonese: | zek3                         | gau2 | soeng6min6      | yau5 | zek3 | gei1hai6sau2 |
|            | CL                           | dog  | (on the) top of | have | CL   | bionic hand  |
|            | "A bionic hand is on a dog." |      |                 |      |      |              |

Turning to  $V_{\text{complex}}$ , as said, they are composed of two classifier-like handshapes with a movement to represent the verb root, as in (14). In an SVO context, it occurred mostly in locative existential and motion directional predicates, except for the DHs of Group 2 and Group 4 who also produced 6 and 2 such tokens in transitive predicates respectively. This  $V_{\text{complex}}$ , which shows some properties of a classifier predicate, may reflect the DHs' initial knowledge of representing the argument relation of the noun referents in an event or a state only. However, it is not associated with abstract morphosyntactic features for referentiality, spatial or subject/object agreement; otherwise, OSV order should occur in their performance, recalling in **Table 9** that OSV only began to occur systematically from Group 2 onwards.

(14)

|        |      |                            |                                  |       |
|--------|------|----------------------------|----------------------------------|-------|
| *HKSL: | DH:  | DOG                        | V <sub>complex</sub> : jump_onto | STONE |
|        | NDH: |                            |                                  |       |
|        |      | "A dog jumps onto a rock." |                                  |       |

To conclude, before attaining native or near-native competence as what the DHs of Group 1 managed to achieve, the DHs of Groups 2, 3 and 4 would initially assume an SVO order with a  $V_{\text{complex}}$  or a  $V_{\text{lexical}}$  for the three types of classifier predicates. These data suggest evidence of language interaction effects in the domains of word order and verb root. The SVO stage

**TABLE 9 |** Production of OSV and SOV orders of DHs and DDs based on adult-like responses.

| Predicate types | Transitive |           | Loc-exist |           | Mot-dir   |           |
|-----------------|------------|-----------|-----------|-----------|-----------|-----------|
|                 | O-S-V (%)  | S-O-V (%) | O-S-V (%) | S-O-V (%) | O-S-V (%) | S-O-V (%) |
| Group 1         | 12 (0.92)  | 0 (0.00)  | 22 (0.92) | 0 (0.00)  | 21 (0.88) | 1 (0.05)  |
| Group 2         | 8 (0.50)   | 1 (0.06)  | 11 (0.46) | 1 (0.04)  | 13 (0.54) | 3 (0.13)  |
| Group 3         | 2 (0.13)   | 2 (0.13)  | 5 (0.21)  | 1 (0.04)  | 3 (0.13)  | 1 (0.04)  |
| Group 4         | 0 (0.00)   | 1 (0.08)  | 1 (0.06)  | 0 (0.00)  | 6 (0.33)  | 2 (0.11)  |
| DD-1            | 6 (1.00)   | 0 (0.00)  | 6 (1.00)  | 0 (0.00)  | 6 (1.00)  | 0 (0.00)  |
| DD-2            | 6 (1.00)   | 0 (0.00)  | 6 (1.00)  | 0 (0.00)  | 6 (1.00)  | 0 (0.00)  |

may stem from crosslinguistic influence from Cantonese and/or the DHs' internal developing HKSL grammar of SVO order with a lexical verb root. However, the observation that more DHs in the senior groups embedded a  $V_{\text{complex}}$  in an SVO or OSV order suggests their increasing morphosyntactic knowledge of this complex predicate, thereby triggering agreement and subsequent syntactic operations like topicalization of the object argument in a construction involving a locative existential predicate.

#### Other mixed structures

For the remaining 29 out of 120 tokens of non-adult-like responses that could not be grouped into a straightforward SVO

equivalent to a localizer in Cantonese. It is followed by the main verb HAVE which is also similar to Cantonese *jau3* with an existential meaning, and the object SCISSORS, hence reflecting an SVO order. The second verb is a  $V_{\text{complex}}$ , comprised of two classifiers to encode a motion directional predicate (i.e., a pair of scissors fall down from the back of a dog). In fact, this string  $SV_{\text{HAVE}} OV_{\text{complex}}$  suggests a derivation based on Cantonese grammar (see 15b); yet, a  $V_{\text{complex}}$  is inserted into the second verb slot at Vocabulary Insertion. Note that 11 out of these 14 tokens of  $V_{\text{complex}}$  displayed an adult-like movement shape to denote a motion directional or a transitive predicate, suggesting that this clause final  $V_{\text{complex}}$  is more like a classifier predicate.

(15a)

|        |      |                                                                                               |                  |      |          |                                               |
|--------|------|-----------------------------------------------------------------------------------------------|------------------|------|----------|-----------------------------------------------|
| *HKSL: | DH:  | DOG                                                                                           | IX <sub>up</sub> | HAVE | SCISSORS | $V_{\text{complex}}$ : a fall <sub>down</sub> |
|        | NDH: |                                                                                               | ----->-----      |      |          |                                               |
|        |      | "There is a pair of scissors on a dog, (the scissors) falls down (from the back of the dog)." |                  |      |          | (DH-G4-1)                                     |

(15b)

|            |                                                          |            |      |           |               |
|------------|----------------------------------------------------------|------------|------|-----------|---------------|
| Cantonese: | gau2                                                     | soeng6min6 | jau3 | gaau3zin2 | dit3 lok6lei4 |
|            | dog                                                      | on top of  | have | scissors  | fall down     |
|            | "A pair of scissors fall down from the back of the dog." |            |      |           |               |

category, we call them "mixed structures" because in some cases we observed mixing of grammatical properties of Cantonese and HKSL in the derivation, in other words, there is the possibility of mixed Numeration (see Table 12). We discarded one token due to our failure of comprehending the string of signs produced by a DH from Group 3.

#### Cantonese-based structure

Seventeen tokens were grouped under this category. Fourteen tokens came from a structure in which the first part of the

(16)

|                 |                                                |              |            |                    |                  |
|-----------------|------------------------------------------------|--------------|------------|--------------------|------------------|
| *HKSL:          | DH:                                            | SCISSOR      | LITTLE^DOG | IX <sub>BACK</sub> | IX <sub>up</sub> |
|                 | NDH:                                           |              |            | --->---            |                  |
| Cantonese base: | gaau3zin2                                      | (hai2)       | zek3       | siu2gau2           | bui3             |
|                 | scissors                                       | (be located) | CL         | small.dog          | back             |
|                 | "There is a pair of scissors on a dog's back." |              |            |                    | up               |
|                 | (DH-G3-1)                                      |              |            |                    |                  |

sentence is contributed by the Cantonese grammar but the final verbal predicate is from HKSL. As shown in (15a), the subject DOG, a location argument, is marked by a pointing sign IX<sub>up</sub>

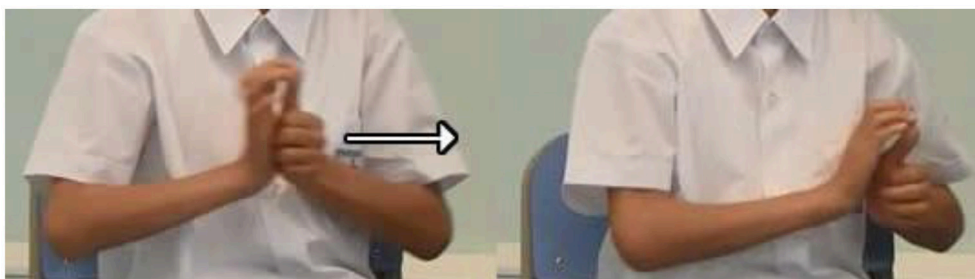
#### Pointing signs as localizers

The second group of data displaying a mixed Numeration came from 3 tokens of utterances produced by the DHs from Groups 3 and 4. The utterances were derived based on the word order of existential predicates in Cantonese but the verb root "be\_located" in HKSL or *hai2* in Cantonese was missing. In place of it, we observed a pointing sign (see 16). *Hai2* in Cantonese is seldom found even in Cantonese-based signing. Therefore, resorting to pointing signs enabled them to encode the locative relation of the two arguments.

Six tokens of locative existential and two tokens of motion directional classifier constructions were nearly adult-like, except that a pointing sign (e.g., IX<sub>up</sub> or

**TABLE 10 |** Non-adult-like occurrences of word order and verb root.

| Erroneous patterns | SVO with a variety of verb roots: 91/120 (76%) |                              |             |                                                     |                             | Total |
|--------------------|------------------------------------------------|------------------------------|-------------|-----------------------------------------------------|-----------------------------|-------|
|                    | S > V <sub>lexical</sub> > O                   | S > V <sub>complex</sub> > O | S > ges > O | S > V <sub>complex</sub> + V <sub>lexical</sub> > O | S > 1-handed Mot-dir CL > O |       |
| Group 1            | 1                                              | 2                            | 0           | 0                                                   | 0                           | 3     |
| Group 2            | 6                                              | 18                           | 0           | 0                                                   | 0                           | 24    |
| Group 3            | 15                                             | 13                           | 2           | 1                                                   | 2                           | 34    |
| Group 4            | 17                                             | 9                            | 2           | 1                                                   | 2                           | 30    |
| Total              | 39                                             | 42                           | 4           | 2                                                   | 4                           | 91    |

**FIGURE 4 |** An example of V<sub>complex</sub> meaning “the cat pushes the backpack”.

IX<sub>back</sub>) was inserted to serve more like a localizer for the location argument, which is redundant in HKSL (see 17).

(17)

|        |      |                               |                          |                   |                                                               |
|--------|------|-------------------------------|--------------------------|-------------------|---------------------------------------------------------------|
| *HKSL: | DH:  | STONE <sub>a</sub>            | BIONIC <sup>^</sup> HAND | aIX <sub>up</sub> | V <sub>complex</sub> : a be <sub>i</sub> located <sub>b</sub> |
|        | NDH: |                               | ----->-----              |                   |                                                               |
|        |      | “A bionic hand is on a rock.” |                          |                   | (DH-G2-1)                                                     |

In summary, the data reveal that Deaf children acquiring classifier constructions in HKSL could converge on the adults’ grammar after 6 to 7 years of exposure. Over time, they could assign a classifier to an R-locus in space using a locative existential predicate, which serves as the grammatical object for the ensuing transitive, locative existential and motion directional predicate for which the classifier on the dominant hand serves as subject. Before attaining this stage of knowledge, we observe evidence of crosslinguistic interaction between Cantonese and HKSL which we will discuss below.

## DISCUSSION

One aim of the current study was to investigate if HKSL-Cantonese DHs, aided or implanted, whose onset of HKSL exposure was not at birth but at age 4 or even as late as age 6 or 7, managed to acquire the complex morphosyntactic properties of classifier constructions. Unlike the Kodas or DDs, their parents are not signers, and the SLCO environment is the only source of HKSL input. The findings show that, despite relatively late exposure to HKSL, these children are able to produce classifier constructions based on an OSV order with R-loci for the classifiers, for subject/object agreement as well as spatial

agreement. In other words, the SLCO environment, designed to provide dual language input, especially HKSL from Deaf teachers and a critical mass of Deaf students on a daily basis, to some

extent offsets the lack of HKSL input in the home environment. In addition to consistent HKSL input and duration of exposure, one other possibility is the Cantonese (and/or written Chinese) input in the SLCO environment, which bolsters bimodal bilingual acquisition and indirectly raises their metalinguistic awareness about differences in word order and verb morphology between Cantonese and HKSL, as well as other properties like the use of space to encode formal grammatical properties like referentiality and agreement (Tang et al., 2015). What we observed among these DHs is the initial adherence to the canonical SVO order and choice of lexical verb root, a property shared by both HKSL and Cantonese. Such a similarity in the morphological property of verbs actually invites crosslinguistic interaction between the two languages, leading to interesting developmental consequences, an issue which we attempt to account for using Language Synthesis.

When predicting effects of crosslinguistic interaction in the current study, we argue that word order and the morphosyntactic properties of the verb are the two domains in which such evidence may be found. The findings reveal that the DHs underwent a protracted SVO stage. During this period, they inserted either a lexical verb (i.e., V<sub>lexical</sub>) or a two-handed verbal sign (i.e., V<sub>complex</sub>) into this SVO structure. Such patterns were quite prominent among the DHs of Groups

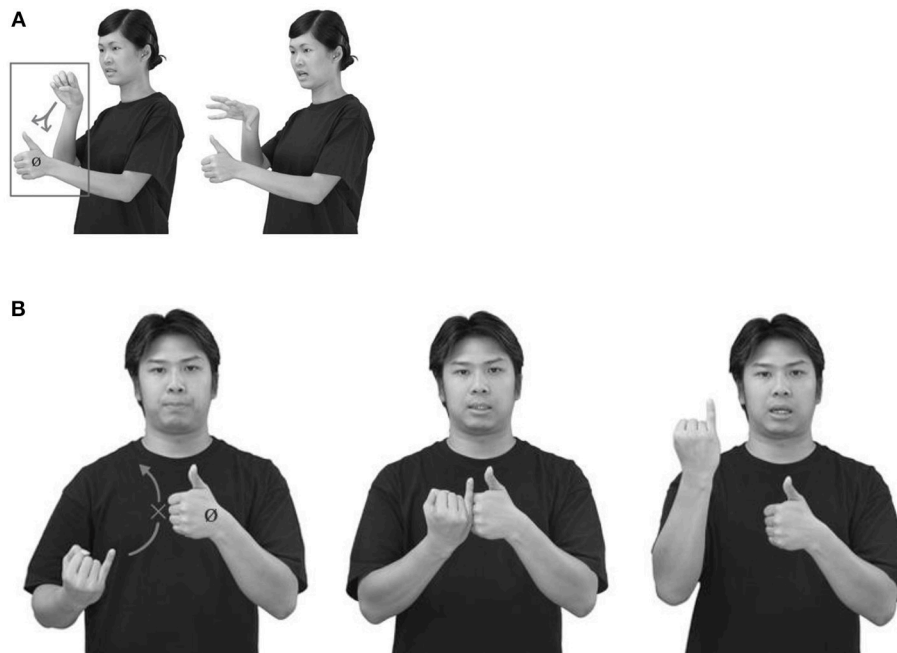


FIGURE 5 | (A) SCOLD, (B) REBEL.

TABLE 11 | Distribution of  $V_{\text{lexical}}$  and  $V_{\text{complex}}$  in a non-adult-like SVO order\*.

| $V_{\text{lexical}}$ and $V_{\text{complex}}$<br>in SVO | Transitive               |                          |        | Loc-exist                |                          |        | Mot-dir                  |                          |        |
|---------------------------------------------------------|--------------------------|--------------------------|--------|--------------------------|--------------------------|--------|--------------------------|--------------------------|--------|
|                                                         | $V_{\text{lexical}}$ (%) | $V_{\text{complex}}$ (%) | Total* | $V_{\text{lexical}}$ (%) | $V_{\text{complex}}$ (%) | Total* | $V_{\text{lexical}}$ (%) | $V_{\text{complex}}$ (%) | Total* |
| Group 1                                                 | 1 (1.00)                 | 0 (0.00)                 | 1      | 0 (0.00)                 | 0 (0.00)                 | 2      | 0 (0.00)                 | 2 (1.00)                 | 2      |
| Group 2                                                 | 1 (0.14)                 | 6 (0.86)                 | 7      | 5 (0.42)                 | 5 (0.42)                 | 12     | 0 (0.00)                 | 7 (0.88)                 | 8      |
| Group 3                                                 | 8 (0.67)                 | 0 (0.00)                 | 12     | 7 (0.39)                 | 3 (0.17)                 | 18     | 0 (0.00)                 | 10 (0.50)                | 20     |
| Group 4                                                 | 3 (0.27)                 | 2 (0.18)                 | 11     | 14 (0.82)                | 2 (0.12)                 | 17     | 0 (0.00)                 | 5 (0.50)                 | 10     |

(\* Total numbers of erroneous responses as denominators).

3 and 4, especially in transitive predicates and locative existential predicates. Examining their non-adult-like tokens, we observed a frequent use of uninflected PUSH for the transitive predicates and HAVE for the locative existential predicates. These verbs have a lexical root which can easily find a translation equivalent in Cantonese such as *toei1* “push” and *jau5* “have.” Take the locative existential predicates as an example, 20 out of 25 non-adult-like tokens adopted HAVE in the predicate. In fact, HAVE in HKSL can be a verb of possession (e.g., KENNY DOG HAVE “Kenny has a dog”), an auxiliary verb encoding perfective aspect of an event (e.g., LAST EVENING IX-1 RUN ONE^HOUR HAVE “Last evening I ran for 1 h”), and as a verb of existence (e.g., HOUSEa IXa DOG HAVE). Clearly, the syntactic position of HAVE is clause-final in the adult’s grammar. However, in all these non-adult-like tokens, HAVE occurs in an SVO structure, which is similar to the existential verb *jau5* in Cantonese, as shown in (7d) above. Therefore, we argue that these children selected the morphosyntactic features of Cantonese initially from List 1, and at Vocabulary Insertion HAVE was selected from HKSL instead. Another piece of evidence for Cantonese influence is

the insertion of a post-nominal pointing sign IX<sub>up</sub> [e.g., in (17) above], which is reminiscent of a Cantonese localizer *soeng6min6* “(on the) top of” to encode the locative relation between two entities (e.g., “a dog on a rock”); however, it is redundant with a locative existential predicate in HKSL. Following Distributed Morphology, we assume it is the assembly of the morphosyntactic features pertaining to Cantonese *jau5* and localizer in the Numeration that determines the syntactic word order, although HAVE from HKSL can be chosen at Vocabulary Insertion. The protracted SVO stage could be a result of DHs not selecting classifier features and locus features in the Numeration initially, as they assumed HKSL verbs are similar to Cantonese which are lexical in nature. As a consequence, the syntactic derivation yields a canonical SVO order and Vocabulary Insertion selects lexical verbs that overlap in Cantonese and HKSL, such as HAVE or uninflected PUSH without subject/object agreement or spatial agreement.

The case of  $V_{\text{complex}}$  is a little complicated. As said, during this protracted SVO stage of development, we also found an increasing number of tokens of two-handed  $V_{\text{complex}}$  alongside

**TABLE 12 |** Occurrences of other mixed structures.

| Error patterns | Other Mixed Structures: 29/120 (24%) |                              |         |       |
|----------------|--------------------------------------|------------------------------|---------|-------|
|                | Cantonese structure                  | Pointing signs as localizers | Discard | Total |
| Group 1        | 0                                    | 2                            | 0       | 2     |
| Group 2        | 1                                    | 2                            | 0       | 3     |
| Group 3        | 9                                    | 7                            | 1       | 17    |
| Group 4        | 7                                    | 0                            | 0       | 7     |
| Total          | 17                                   | 11                           | 1       | 29    |

the  $V_{\text{lexical}}$ . Arguably, features for classifiers are selected and spelt out as a two-handed  $V_{\text{complex}}$ , leading to a change of the morphological structure of the verb. However, other features, especially the locus feature for spelling out the R-loci for spatial agreement, are not selected initially. In other words, without the locus feature, no object agreement node is projected above little  $\nu P$ . At PF, it is the lack of R-loci for spatial agreement rather than classifier agreement that marks the  $V_{\text{complex}}$  distinct from those observed in the classifier predicates produced by the adult signers. Subsequent acquisition of classifier predicates, in particular the selection of locus feature for spatial agreement will lead to a further reanalysis of the morphological status of  $V_{\text{complex}}$ . Such reanalysis triggers decomposition of the two-handed signs and the copying of classifier and locus features to different agreement nodes for structural agreement with the noun arguments at the specifier positions. Furthermore, the development of pragmatic knowledge involved in the signing discourse by these children also led to the object being introduced independently by a locative classifier predicate or probably through a movement operation to the left periphery. We leave this part of the analysis for future research.

Although our research is not particularly geared toward analyzing code blending, some of our data like those discussed above resemble what (Branchini and Donati, 2016) would refer to as Type 1 (i.e., Cantonese or HKSL-based) or Type 3 (i.e., mixed) structures. As for Type 3, as said, we found some mixed use of HKSL and Cantonese grammars. For instance, an additional verb of existence HAVE reflecting the Cantonese grammar is adopted to introduce a noun referent (usually the Figure), which is followed by  $V_{\text{complex}}$  at the clause-final position (see descriptions above under “Cantonese-based Structure”). In fact, 14 such erroneous mixed structures were produced by the DHs of Group 3 (i.e., 3 tokens for locative predicates and 4 tokens for motion directional predicates) and Group 4 (i.e., 1 token for locative predicates, 3 tokens for motion directional predicates, and 3 tokens for transitive predicates). In Cantonese, *jau5* “have”+NP introduces a theme argument whereas in HKSL it is introduced by a locative classifier predicate or some localization strategies. Therefore, what we believe to be evidence of a mixed structure came from the erroneous productions like (15a). Although the DHs adopted a Cantonese SVO structure, they attached a clause-final classifier predicate after the object. In other words, the lack of a direct translation equivalent for a lexical locative existential *hai2* “be located” to be signed in

such a way that it becomes head final simply goes against the Cantonese grammar whose verbs are consistently head initial. As for the motion directional predicates, if the DHs followed the Cantonese grammar entirely, they had to produce three independent signs (i.e., VVV) due to serial verb constructions which uniquely occur in Cantonese but not in HKSL so far as motion directional predicates are concerned. That the DHs were in a test condition for HKSL production encouraged them to switch to the HKSL structure and choose a  $V_{\text{complex}}$  in some spatial configuration with a path movement with two endpoints to encode the source and the goal of the predicate. This finding also gives us some clues as to why they performed better on motion directional predicates than other predicates in the current study.

To sum up, this study reveals that Deaf children undergoing bimodal bilingual acquisition showed co-activation of the two grammars in the Numeration, during which they assumed knowledge of word orders available from the two languages, and the so called “mixing” occurred primarily in the verbal domain in their outputs. Among all the features they need to acquire for classifier constructions, the results show that locus features were acquired last in the process.

## CONCLUSION

Although earlier studies showed that classifier predicates may emerge more or less the same time as agreement verbs, full mastery was consistently reported to be late, owing to their morphosyntactic complexity. The current study revealed that consistent HKSL input over time could lead to convergence on the adult’s grammar, despite a lack of early exposure to the language since birth. Where the home environment does not facilitate sign language acquisition, the school environment with consistent HKSL input from Deaf adults and Deaf peers becomes crucial for supporting the DHs’ HKSL development. This echoes the findings from some previous studies that consistent sign language exposure in schools facilitates Deaf children’s sign language development (Henner et al., 2016 on ASL; Tomasuolo et al., 2010 on LIS). As the SLCO learning environment is newly established and the size of the sample is quite small, more acquisition research with Deaf children from this environment is necessary in order to verify if it positively impacts their sign language acquisition.

At the theoretical level, this study attempts to apply Language Synthesis to account for the acquisition phenomena. The data confirm that Numeration from List 1 and Vocabulary Insertion are the two domains in which one may examine crosslinguistic interaction. This kind of research is still preliminary. In future, other structures which show typological differences or even similarities may be incorporated into the investigation.

## AUTHOR CONTRIBUTIONS

GT main and corresponding author, oversee the design, outline of the paper, theoretical framework, data collection and processing,

as well as writing up and editing the paper. JL processing the data and writing up the results section of the paper.

## FUNDING

This current study is part of a large-scale deaf education program funded by donations from the Hong Kong Jockey Club Charities Trust, Lee Hysan Foundation, and Fu Tak Iam Foundation. The linguistic analysis of Hong Kong Sign Language is supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK

450513) and CUHK-Faculty of Arts, The Publication Subvention Fund awarded to GT.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the colleagues, the deaf children and their parents of the Sign Bilingualism and Co-enrollment in Education Programme. Also, the support of the Deaf researchers in conducting the experiments and annotating the data should not go unnoticed.

## REFERENCES

- Baker, A. E. (2016). Incongruent grammar: can the model cope? *Ling. Appr. Bilingualism* 6, 760–762. doi: 10.1075/lab.6.6.03bak
- Benedicto, E., and Brentari, D. (2004). Where did all the arguments go?: argument-changing properties of classifiers in ASL. *Natl. Lang. Ling. Theor.* 22, 743–810. doi: 10.1007/s11049-003-4698-2
- Branchini, C., and Donati, C. (2016). Assessing lexicalism through bimodal eyes. *Glossa J. General Ling.* 1, 1–30. doi: 10.5334/gjgl.29
- Chen Pichler, D. (2001). *Word Order Variability and Acquisition in American Sign Language*. Ph.D. dissertation, University of Connecticut.
- Cogill-Koez, D. (2000). Signed language classifier predicates: linguistic structures or schematic visual representation? *Sign Lang. Ling.* 3, 153–207. doi: 10.1075/sll.3.2.03cog
- Cormier, K., Quinto-Pozos, D., Sevcikova, Z., and Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: comparing depicting constructions and viewpoint gestures. *Lang. Commun.* 32, 329–348. doi: 10.1016/j.langcom.2012.09.004
- Crasborn, O., and Sloetjes, H. (2008). “Enhanced ELAN functionality for sign language corpora,” in *Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation*, 39–43. Available online at: <http://www.lrec-conf.org/proceedings/lrec2008>
- Donati, C., and Branchini, C. (2013). “Challenging linearization: simultaneous mixing in the production of bimodal bilinguals,” in *Challenges to Linearization*, eds T. Biberauer and I. Roberts (Berlin: Mouton De Gruyter), 93–128.
- Emmorey, K., Borinstein, H. B., Thompson, R., and Gollan, T. H. (2008). Bimodal bilingualism. *Bilingual. Lang. Cogn.* 11, 43–61. doi: 10.1017/S1366728907003203
- Fung, H.-M. C., and Tang, G. (2017). Code-blending of functional heads in Hong Kong Sign Language and Cantonese: a case study. *Bilingual. Lang. Cogn.* 19, 754–781. doi: 10.1017/S1366728915000747
- Glück, S., and Pfau, R. (1998). “On classifying classification as a class of inflection in German Sign Language,” in *Proceedings of ConSOLE VI*, eds T. Cambier-Langeveld, A. Lipták, and M. Redford (Leiden: SOLE), 59–74.
- Glück, S., and Pfau, R. (1999). “A Distributed Morphology account of verbal inflection in German Sign Language,” in *Proceedings of ConSOLE VII*, eds T. Cambier-Langeveld, A. Lipták, M. Redford, and van der E. J. Torre (Leiden: SOLE), 65–80.
- Harley, H. (2014). On the identity of roots. *Theor. Ling.* 40, 1–53. doi: 10.1515/tl-2014-0010
- Henner, J., Caldwell-Harris, C. L., Novogrodsky, R., and Hoffmeister, R. (2016). American Sign Language syntax and analogical reasoning skills are influenced by early acquisition and age of entry to signing schools for the Deaf. *Front. Psychol.* 7:1982. doi: 10.3389/fpsyg.2016.01982
- Huang, C. T.-J. (1990). On ‘be’ and ‘have’ in Chinese. *Bull. Inst. Hist. Philol. Acad. Sin.* 59, 43–64.
- Huang, C. T.-J. (2009). Lexical decomposition, silent categories, and the Localizer Phrase. *Yuyanxue Luncong* 39, 86–122.
- Hulk, A., and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingual. Lang. Cogn.* 3, 227–244. doi: 10.1017/S1366728900000353
- Koulidobrova, H. (2012). *When the Quiet Surfaces: ‘Transfer’ of Argument Omission in the Speech of ASL-English Bilinguals*. Ph.D. dissertation, University of Connecticut.
- Koulidobrova, H. (2016). Language interaction effects in bimodal bilingualism: argument omission in the languages of hearing ASL-English bilinguals. *Ling. Appr. Bilingual.* 75, 583–613. doi: 10.1075/lab.13047.kou
- Lam, W. S. (2009). *Early Phrase Structure in Hong Kong Sign Language: A Case Study*. Ph.D. dissertation, The Chinese University of Hong Kong.
- Lillo-Martin, D. (1999). “Modality effects and modularity in language acquisition: the acquisition of American Sign Language,” in *Handbook of Language Acquisition*, eds W. Ritchie and T. Bhatia (San Diego, CA: Academic Press), 531–567.
- Lillo-Martin, D., and Berk, S. (2003). “Acquisition of constituent order under delayed language exposure,” in *Proceedings from 27<sup>th</sup> Annual Boston University Conference on Language Development*, eds B. Beachley, A. Brown, and F. Conlin (Somerville, MA: Cascadilla Press), 484–495.
- Lillo-Martin, D., de Quadros, R. M., Chen Pichler, D., and Fieldsteel, Z. (2014). Language choice in bimodal bilingual development. *Front. Psychol.* 5:1163. doi: 10.3389/fpsyg.2014.01163
- Lillo-Martin, D., de Quadros, R. M., Koulidobrova, H., and Chen Pichler, D. (2010). “Bimodal bilingual cross-language influence in unexpected domains,” in *Language Acquisition and Development: Proceedings of GALA 2009*, eds J. Costa, A. Castro, M. Lobo, and F. Pratas (Newcastle upon Tyne: Cambridge Scholars Press), 264–275.
- Lillo-Martin, D., Koulidobrova, H., de Quadros, R. M., and Chen Pichler, D. (2012). “Bilingual language synthesis: evidence from wh-questions in bimodal bilinguals,” in *Proceedings of the 36th Annual Boston University Conference on Language Development*, eds A. K. Biller, E. Y. Chung, and A. E. Kimball (Somerville, MA: Cascadilla Press), 302–314.
- Lillo-Martin, D., de Quadros, R. M., and Chen Pichler, D. (2016). The development of bimodal bilingualism: implications for linguistics theory. *Ling. Appr. Bilingual.* 6, 719–755. doi: 10.1075/lab.6.6.01lil
- Lindert, R. (2001). *Hearing Families with Deaf Children: Linguistic and Communicative Aspects of American Sign Language Development*. Ph.D. dissertation, University of California, Berkeley.
- MacSwan, J. (2000). The architecture of the bilingual language faculty: evidence from intrasentential code switching. *Bilingual. Lang. Cogn.* 3, 37–54. doi: 10.1017/S1366728900000122
- MacSwan, J. (2005). Codeswitching and generative grammar: a critique of the MLF model and some remarks on “modified minimalism.” *Bilingual. Lang. Cogn.* 8, 1–21. doi: 10.1017/S1366728904002068
- Meisel, J. M. (2011). *First and Second Language Acquisition: Parallels and Differences*. Cambridge: Cambridge University Press.
- Newport, E., and Meier, R. (1985). “The acquisition of American Sign Language,” in *The Crosslinguistic Study of Language Acquisition*, ed D. Slobin (Mahwah, NJ: Lawrence Erlbaum), 881–938.
- Ng, H. Y. I. (2014). *The Construction and Validation of the Cantonese Spoken Word Recognition Test (CanSWORT) to Measure Word Recognition Ability of Cantonese-speaking Population*. Ph.D. dissertation, The Chinese University of Hong Kong.
- Palmer, J. L. (2015). *Bimodal Bilingual Word Order Acquisition*. Ph.D. dissertation, Gallaudet University.

- Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., and Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: implications for the mechanisms underlying early bilingual language acquisition. *J. Child Lang.* 28, 453–496. doi: 10.1017/S0305000901004718
- Pfau, R., and Aboh, E. (2012). On the syntax of spatial adpositions in sign languages. *MIT Work. Papers Ling.* 65, 83–104.
- Schembri, A. (2003). “Rethinking ‘classifiers’ in signed languages,” in *Perspectives on Classifier Constructions in Sign Languages*, ed K. Emmorey (Mahwah, NJ: Psychology Press), 3–34.
- Schick, B. (1990). “The effects of morphosyntactic structure on the acquisition of classifier predicates in ASL,” in *Sign language research: Theoretical issues*, ed C. Lucas (Washington, DC: Gallaudet University Press), 358–374.
- Supalla, T. (1982). *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. Ph.D. dissertation, University of California, San Diego, CA.
- Supalla, T. (1986). “Classifier system in American Sign Language,” in *Noun Classes and Categorization*, ed C. Craig (Amsterdam/Philadelphia: John Benjamins Publishing Company), 181–214.
- Sze, F. (2000). “Word order of Hong Kong Sign Language,” in *Cross-linguistic Perspectives in Sign Language Research*, eds A. Baker, B. van den Bogaerde, and O. Crasborn (Hamburg: Signum), 163–192.
- T’sou, B., Lee, T., Tung, P., Chan, A., Man, Y., and To, C. (2006). *Hong Kong Cantonese Oral Language Assessment Scale*. Hong Kong: City University of Hong Kong Press.
- Talmy, L. (2000). *Toward a cognitive semantics. Vol. 1: Concept Structuring Systems*. Cambridge, Mass: MIT Press.
- Tang, G., Lam, S., and Yiu, K.-M. C. (2014). “Language development of deaf children in a sign bilingual and co-enrollment environment,” in *Bilingualism and Bilingual Deaf Education*, eds M. Marschark, G. Tang, and H. Knoors (New York, NY: Oxford University Press), 313–341.
- Tang, G., Sze, F., and Lam, S. (2007). “Acquisition of simultaneous constructions by deaf children of Hong Kong Sign Language,” in *Simultaneity in Signed Languages: Form and Function*, eds M. Vermeerbergen, L. Leeson, and O. Crasborn (Amsterdam: Benjamins), 283–316.
- Tang, G., Yiu, C., and Lam, S. (2015). “Awareness of Hong Kong sign language and manually coded chinese by deaf students,” in *Educating Deaf Learners: Creating a Global Evidence Base*, eds H. Knoors and M. Marschark (New York, NY: Oxford University Press), 117–148.
- Tomasuolo, E., Fellini, L., Renzo, A. D., and Volterra, V. (2010). Assessing lexical production in deaf signing children with the Boston Naming Test. *Lang. Interact. Acquisit.* 1, 110–128. doi: 10.1075/lia.1.1.07tom
- van den Bogaerde, B., and Baker, A. (2005). Code mixing in mother-child interaction in deaf families. *Sign Lang. Linguist.*, 8, 153–176. doi: 10.1075/sll.8.1.08bog
- Yuen, K. C. P., Ng, I. H. Y., Luk, B. P. K., Chan, S. K. W., Chan, S. C. S., Kwok, I. C. L., et al. (2008). The development of Cantonese Lexical Neighborhood Test: a pilot study. *Int. J. Pediatric Otorhinolaryngol.* 72, 1121–1129. doi: 10.1016/j.ijporl.2008.03.025
- Zwitserslood, I. (2003). *Classifying Hand Configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Ph.D. dissertation, Utrecht University.
- Zwitserslood, I. (2008). “Morphology below the level of the sign-frozen forms and classifier predicates,” in *Signs of the Time: Selected Papers from TISLR 2004*, ed J. Quer (Hamburg: Signum), 251–272.

**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 © 2018 Tang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Is Morphological Awareness a Relevant Predictor of Reading Fluency and Comprehension? New Evidence From Italian Monolingual and Arabic-Italian Bilingual Children

Mirta Vernice<sup>1\*</sup> and Elena Pagliarini<sup>2</sup>

<sup>1</sup> Department of Psychology, Università degli Studi di Milano-Bicocca, Milan, Italy, <sup>2</sup> Center for Brain and Cognition (CBC), Departament de Tecnologies de la Informació i les Comunicacions (DTIC), Universitat Pompeu Fabra, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Maria Vender,  
University of Verona, Italy  
Jana Hasenäcker,  
Scuola Internazionale Superiore di  
Studi Avanzati (SISSA), Italy

### \*Correspondence:

Mirta Vernice  
mirta.vernice@unimib.it

### Specialty section:

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

**Received:** 10 May 2017

**Accepted:** 06 March 2018

**Published:** 23 March 2018

### Citation:

Vernice M and Pagliarini E (2018) Is Morphological Awareness a Relevant Predictor of Reading Fluency and Comprehension? New Evidence From Italian Monolingual and Arabic-Italian Bilingual Children. *Front. Commun.* 3:11. doi: 10.3389/fcomm.2018.00011

In this study, we examined the contribution of morphological awareness to reading competence in a group of Italian L1 and Arabic-Italian early L2 children, i.e., exposed to Italian before 3 years of age. Children from first to fifth grade (age range: 6–11 years old) were tested on a range of morphological awareness and lexical tasks. Reading ability was tested through standardized tests of reading fluency and comprehension. Results showed that L1 children outperformed L2 on every measure of morphological awareness, as well as on reading tests. Regression analyses revealed that morphological awareness contributed to a different extent to reading ability across groups. Accuracy in the morphological awareness tasks was a significant predictor of word (and non-word) reading fluency in L1 and L2 first and second graders, while only in L1 third to fifth graders, response times and accuracy to a morphological awareness task explained a unique amount of variance in reading comprehension. Our results highlight the critical role of morphological processing in reading efficiency and suggest that a training inspired by morphological awareness may improve reading skills also in bilingual students.

**Keywords:** reading achievement, morphological awareness, derivational morphology, reading comprehension, L2 children, reading in L2 children

## INTRODUCTION

In recent years, research has often investigated the linguistic underpinnings of reading development, highlighting the role of phonological skills and vocabulary as significant predictors of literacy achievement (Gathercole and Baddeley, 1989, 1993; Baddeley et al., 1998). From this standpoint, adequate phonological skills are a prerequisite for the development of optimal phonological representations of words in the mental lexicon (Fowler, 1991), and, as a consequence, of reading development (e.g., Baddeley et al., 1998). This study goes further by exploring another possible linguistic predictor of reading: morphological awareness, i.e., the consciousness of how complex words are made up of smaller units and the ability to manipulate those units to generate a new word (Carlisle, 2000; Kuo and Anderson, 2006). Since evidence based on young or impaired readers suggest that they benefit from a morphological parsing strategy in reading (Casalis et al., 2004; in Italian, Burani et al., 2008; Angelelli et al., 2014), it appears important to explore the role that morphological awareness plays in

reading development both in monolingual and bilingual populations. In this study, we tested the performance of monolingual and bilingual reading learners on a range of morphological awareness tasks and reading tests. By doing so, we aimed at gaining a clearer understanding of the relationship between morphological awareness and reading achievement.

It is known that reading development is a complex cognitive and linguistic process that involves several underlying cognitive abilities, such as phonological awareness, vocabulary, and grammatical skills (cf. Nagy and Townsend, 2012). According to the literature, however, beyond phonological awareness and orthographic competence, also morphological awareness might be considered an additional predictor not only of word reading fluency (Fowler and Liberman, 1995; Carlisle and Katz, 2006; Roman et al., 2009) but, most importantly, of reading comprehension (Deacon and Kirby, 2004; Nagy et al., 2006; Tong et al., 2011). Therefore, morphological awareness seems to be a potential, interesting, underlying ability that might significantly contribute to the study of reading development (Carlisle, 1995; Deacon and Kirby, 2004; Roman et al., 2009).

Let us now briefly summarize what is generally intended by morphological awareness. Morphological awareness refers to the metalinguistic consciousness that words are constituted of individual units (i.e., morphemes) which can be analyzed and manipulated in various ways (Carlisle, 1995; Derwing et al., 1995; Kuo and Anderson, 2006). Roughly, there are three types of morphological operations that allow the creation of new word forms: inflection, derivation, or compounding. Inflectional processes allow the modification of grammatical aspects of the word such as number, gender, and tense (e.g., boy-s; open-ed), while derivational operations generate new words by changing, in some cases, the meaning of the root (e.g., easy; un-easy) and usually (but not necessarily) its grammatical category (e.g., “strong-ly” is the adverbial form of the adjective “strong”; however “farmer” is a noun that derives from the noun “farm” and it is used to refer to the person who runs a farm). Compounding mechanisms, on the other hand, generate new words, combining two autonomous lexical units (dish and washer) into a new word (e.g., dishwasher).

This study will focus on derivational morphology. From a developmental perspective, derivational formation might require a deeper knowledge of the complex association between morphemes and their meanings. That is, morphological awareness of derived words’ composition involves knowledge of the semantic underpinnings of prefixes (e.g., the un- in unpleasant, with the prefix involving a meaning of negation) and suffixes (e.g., the -er in sing-er, with the suffix -er denoting agentivity). For this reason, while inflectional morphology tends to develop relatively early, derivational morphology knowledge continues to develop throughout school years (Casalis and Louis-Alexandre, 2000).

It should be noted that a difficulty in processing derivationally complex words, whose meaning is often unfamiliar to students, might hinder reading and comprehension of a new text. Therefore, investigating to what extent morphological awareness might scaffold children’s ability to read and comprehend complex unfamiliar words might have important practical implications. This study aims to answer this question by focusing on the contribution of derivational morphological knowledge on reading achievement.

Let us now focus on the specific role that morphological awareness might play in reading development in the course of literacy acquisition (Tong et al., 2011), by distinguishing the specific contribution it exerts on beginning vs. competent readers. In young readers (i.e., according to the Italian school system, first and second graders, that are still learning to read and that have not automatized the reading process yet), morphological decomposition ability might allow parsing the word by analyzing it in smaller units (Rispen et al., 2008). There is a bulk of evidence indicating that young Italian readers tend to implicitly parse a word in smaller units (morphemes; Burani et al., 2008) to facilitate the reading process. According to these studies, words with a morphological structure (e.g., cass-iere, “cashier”) were read faster than simple words (e.g., cammello, “camel”) matched for length and frequency. Interestingly, morphological parsing speeded up reading times only in second graders and in children with dyslexia, but not in older skilled children (Burani, 2010; Marcolini et al., 2011; Angelelli, 2010; Angelelli, 2017). The authors concluded that children acquiring a transparent orthography such as Italian exploit morpheme-based reading and spelling to face difficulties in reading long unfamiliar words. Even though the previous studies do not refer to an explicit measure of morphological awareness, they showed that (implicit) morphological processing enhances reading performance in Italian young readers by facilitating the parsing of a complex word through decomposition.

Note that such findings are in line with the claim that in lexical access, readers are sensitive to the internal morphological representation of orthographically transparent (Baayen et al., 1997; Marslen-Wilson and Tyler, 1997) as well as opaque words (i.e., where morpheme meanings are inconsistent with word meaning; Rastle et al., 2004). For instance, a series of masked priming studies indicated that “corner,” which can be inappropriately segmented as corn + er (though a corner is not someone who “corns”) facilitated word recognition of CORN as pairs like dealer-DEAL, where primes and targets entertain a genuine morphological relationship (Crepaldi et al., 2010). That is, according to these studies, lexical elaboration may be sensitive to the internal morphological representation but not to the semantics of the morphemes (Rastle et al., 2004).

Additional evidence (Amenta et al., 2015) indicates that early morphological analysis in lexical access is sensitive to the semantic representations of the individual morphemes even in opaque words. That is, reading the Italian word *bottone* (button), at a very early processing stage, would automatically activate the representation “bott-one” (loud thud), significantly slowing down first fixations of this morphologically opaque word (Amenta et al., 2015). Consequently, one might claim that when we access a word such as *secchione* (nerd) we also process its underlying surface morphological structure (secchione; big bucket).

Overall, such findings suggest that, once the decoding process in reading is automatized, morphological analysis and decomposition might support the ability to make lexical inferences about the internal structure of complex words. According to the literature, such competence might facilitate the comprehension of unfamiliar words (Carlisle, 2007), and, as a consequence, the comprehension of the whole text. That is, decomposition processes (of derived

words) would facilitate the extraction of semantic and syntactic information that supports reading comprehension of connected text (Kieffer et al., 2013). Thus, morphological awareness appears as a critical prerequisite of lexical analysis not only at word-level but also at text-level promoting lexical inference along the course of literacy acquisition.

Note that the previous studies were conducted on monolingual children. But what do we know about the development of a metalinguistic ability such as morphological awareness in bilingual children? Recent studies on this topic provided compelling evidence, suggesting a strong bilingual advantage in the development of metalinguistic awareness (e.g., Bialystok et al., 2014). For instance, Bialystok observed that English speaking children who entered a French immersion program at school, outperformed their peers (enrolled in a monolingual program) when undertaking a series of metalinguistic tasks, among which was the well-known Berko's Wug Test (Berko, 1958), which is based on inflectional morphology and proposed in the L1 of the children, namely English. The authors conclude that after only 2 years in an immersion education program, children showed some of the metalinguistic advantages generally associated with fully bilingual children (Bialystok et al., 2008). Another recent study (Kuo et al., 2017) demonstrated that L1 Spanish and L1 English children enrolled in a dual (English-Spanish) program showed better morphological derivational awareness both in English and Spanish in comparison with their peers in general education. The development of metalinguistic skills, with a specific focus in derivational morphology, in bilingual children appeared to have been enhanced by cross-language transfer of cognate words, i.e., words that show an overlap in form and meaning across languages as well as by an increased sensitivity to structural language features. As the authors note, indeed, many low-frequency academic words in English derive from the same stem of high-frequency words in Spanish (e.g., English *tranquil* and Spanish *tranquilo*; Proctor and Mo, 2009).

Previous studies have tried to address a further issue, namely, the relationship between morphological awareness and reading fluency and comprehension in L2 children (Goodwin et al., 2011; Ramirez et al., 2011; Kieffer et al., 2013). A study conducted on Arabic-English children demonstrated that morphological awareness exerts a cross-linguistic influence on reading fluency: for instance, Arabic morphological awareness predicts English word reading (Saiegh-Haddad and Geva, 2008). However, additional studies confirm that the correlation between morphological awareness and reading fluency in L1 and L2 appears to be strongly mediated by a child's phonological awareness and lexical abilities in both languages spoken (Goodwin et al., 2011; Ramirez et al., 2011). Accordingly, Kieffer et al. (2013) showed that morphological awareness predicted reading comprehension but only when controlling for lexical competence. Overall, taking together the above-mentioned findings, one might conclude that morphological awareness might appear enhanced in bilingual speakers, but its role in reading development is strictly linked to the child's lexical knowledge.

This study was designed to test the contribution of morphological awareness to reading fluency and comprehension in monolingual and bilingual children. The first aim of this study was

to disentangle the predictive ability of morphological awareness in the development of reading competence on 41 L1 Italian children whose age ranged from 6 to 11. The second aim was to compare the morphological awareness of 12 Arabic-Italian speakers (age range 6–11 years of age) with 12 age-matched L1 Italian speakers. By doing so, we meant to investigate to what extent this competence contributed to reading achievement in L1 and L2 learners. Given the fact that Italian L2 readers demonstrate slow and often inaccurate reading performance (e.g., Murineddu et al., 2006), investigating the effects of morphological awareness on reading ability in this population might provide theoretical and practical implications to improve their academic performance.

To sum up, we propose the following predictions, which might apply to both L1 and L2 children. First, if a child still relies on word decoding to read fluently, then morphological awareness should influence reading fluency at word-level (e.g., Burani et al., 2008). Second, if a child has fully automatized decoding, she or he must be able to access a lexical unit fully defined from an orthographic, lexical, and semantic perspective; then morphological knowledge should support higher-order skills such as reading comprehension, enabling readers to make inferences about the meaning of morphologically complex words. In such a case, morphological awareness, together with other factors such as lexical knowledge, should affect reading comprehension at text-level.

Note that such predictions might apply to L1 and L2 children depending on their inherent reading competence (Bellocchi and Genesee, 2012). It is important to remember that according to previous studies based on Italian L2 speakers (Murineddu et al., 2006), L2 children might show a delay in automatizing reading skills resulting in a profile of learning difficulty. We expect therefore that morphological awareness could play a different role in L1 vs. L2 groups, according to their stage of reading development.

## MATERIALS AND METHODS

### Participants

A total of 53 children who attended a local public primary school in the Milan area, Italy, participated in this study. Participants ranged in age from 6;1 (years;months) to 10;11 (mean age = 8;2, SD = 1;3) and were enrolled in first through fifth grade, according to the Italian school system (first grade: 6–7 years old; second grade: 7–8; third: 8–9; fourth: 9–10; fifth: 10–11 years old). Children were divided into two groups: Arabic-Italian speaking bilingual children (L2,  $n = 12$ ; 5M; age range: 6;1–10;11; mean age = 7;7, SD = 1;4) and monolingual Italian (L1,  $n = 41$ ; 16M; age range: 6;2–10;11; mean age = 8;5, SD = 1;3). L1 children were subsequently divided into two groups according to the class they belonged to: beginning readers *L1* (21 children; 8M; age range: 6;2–8;0; mean age = 7;6, SD = 0;3), involving only first and second graders, and competent readers *L1* (20 children; 8M; age range: 7;11–10;11; mean age = 9;7, SD = 1;3), involving third to fifth graders. Beginning and competent L1 readers significantly differed in chronological age ( $t = 6.89, p < 0.001$ ).

The choice to select groups with respect to grades was grounded on the fact that, in Italian, a child's ability to read is

known to become automatized and effortless from the third grade onward (Zoccolotti et al., 2009). That is, accuracy levels for word reading reach ceiling by third grade, with reading speed improving more slowly since then (Tressoldi et al., 2001). Further evidence relies on the fact that reading fluency of (low-frequency) words after third grade show a significant increase with respect to non-words reading (Orsolini et al., 2006). Additional data indicate that in Italian first and second graders, reading skills appear to be predicted to a great extent only by phonological awareness and RAN, while from third grade on reading competence is no longer influenced by phonological skills, but by vocabulary, RAN, verbal memory (digit span), and visuospatial attention (Tobia and Marzocchi, 2014). The authors propose that according to the level of reading automation, readers might selectively activate different cognitive mechanisms.

Regarding the L2 group, all children could be regarded as early bilingual (eight of them were born in Italy, four of them arrived before 2 years of age; cf. Kovelman et al., 2008). By early bilingual, we refer to children who were exposed to a (minority) language (i.e., Arabic) from birth as a first language (L1) and began to learn the L2 (Italian) after they had been enrolled in Italian-only kindergarten at age 3. For each of them, we collected information about their exposure in months to Italian by means of a simple questionnaire that was completed by the parents. By doing so, we were able to test whether (traditional) length of exposure could affect children's performance to morphological tasks and/or reading tests. In general, children had an average of 5;6 years of exposure ( $SD = 1;1$ ) to L2 Italian as a curricular language in pre-school and school.

To compare their performance with monolingual peers, 12 L1 children (out of the total of 41 children) were matched as close as possible to the bilingual participants (L2 group) with respect to age ( $\pm 2$  months) and gender. The two groups did not differ with respect to age (months) ( $t = -0.066$ ,  $p = 0.948$ ).

All the children came from a middle-low SES background, as emerged from a questionnaire that all the parents had to fill in, indicating their job and educational level. In the L1 group, most of the parents had a high school or a university degree; in the L2 group, at least one parent in each couple attained a high school degree in the home country.

To be included in both (L1 and L2) groups, children had to meet a number of criteria. First, none of them had to report a cognitive, neurological, sensorial disability. Second, none of them had to be identified as needing special educational support (according to teachers' reports). Written informed consent was obtained from the parents of all participating children in compliance with the guidelines of our Ethical Committee. The protocol was approved by the Ethical Committee of the University of Milano-Bicocca (IBR: no. 20974/13).

## Materials

To address our research question, participants in both groups took part in three experimental tasks of morphological awareness and one of lexical ability. In addition, they were administered a battery of standardized tests of reading fluency (word- and text-level) and comprehension.

## Morphological Awareness and Lexical Tasks

To study morphological awareness and lexical competence, we created three computerized tasks that investigated morphological awareness both in comprehension and in production. The tasks were presented on a laptop computer using E-Prime software 1.2 (Schneider, 2002), and were designed to be individually executed by the child under the supervision of the experimenter. Participants received oral and written instructions. For Tasks 2 and 3, which involved an oral response of the child, answers were recorded and scored off line.

Task 1 tackles the comprehension of nominal derivational morphology. Children were simultaneously orally and visually presented with pairs of words and were asked to distinguish those that were morphologically (as well as lexically and semantically) related (as in "anello-anellino" *ring-little ring*) or not (as in "burro-burrone" *butter-ravine*), by pressing the YES or the NO button on the keyboard. Note that in this task (as well as and in the production Task 3, see below) we opted for a simultaneous visual and oral presentation. We did so because a visual-only presentation would have been deeply affected by the reading skills of participants. An oral-only presentation could be possibly affected by lack of listening comprehension of the verbal string (possible in L2 children, but also plausible in L1 ones). Therefore, longer RTs could be caused on the one hand by struggle with reading, on the other by a problem in listening comprehension. By simultaneously presenting both orally and visually our stimuli, we were able to control these possible sources of bias.

Both accuracy scores and RT measures were obtained for each trial. Participants were provided with 4 practice items, which were followed by the 32 experimental items (for a list of the experimental items, see Appendix A).

Task 2 was a production task: participants had to recognize a morphological (lexical-semantic) relationship between the object visually and orally presented in the first picture ("campana" *bell*) and the target picture that the child had to name (i.e., "campanile" *bell tower*). The test comprised 16 experimental trials in addition to 6 practice items. Each trial involved a picture and its verbal description and a morphologically related target picture that had to be named out loud. The accuracy of the verbal responses was assessed as dependent variable. In this task, in contrast to Task 1, children were exposed to a concurrent presentation of oral and pictured version of items. Also in this case, we opted for a simultaneous presentation to facilitate comprehension of the experimental items.

Task 3 was a production task: children were orally and visually presented with a sentence that could involve or not a lexical mistake (i.e., "\*Gianni ha mangiato un arrosto di tacco" \**Gianni ate roasted heel* or "Silvia coltiva la salvietta" \**Silvia cultivates the towel*). They had to detect the anomaly (if present), and to correct it by generating an appropriate (non-morphologically and non-semantically, but phonologically related) word (i.e., "tacchino," *turkey*, instead of "tacco" *heel* or "salvia," *sage*, instead of "salvietta" *towel*). Note that, as shown in our example, the incorrect lexical item (tacco) and the target (tacchino) were semantically independent, but involved a surface morphological relationship, i.e., words were both made up of a pseudo-stem, which was shared

with the targets, and of a pseudo-suffix. Therefore, to correct the sentence, participants had to generate a new word by adding (as “-ino” in “tacchino”) or deleting (as -etta in “salvietta”) a pseudo-suffix. Note that to accomplish the task (i.e., correct the sentence), children could select other semantic plausible but phonologically unrelated words, i.e., selecting an appropriate but non-target word (i.e., “pianta” vegetable). Therefore, if participants were able to identify the target word, it would suggest that they relied on the decomposition of the morphological structure of the opaque word (“salvi-etta”) to access the target (“salvia”) (cf. Amenta et al., 2015).

In each experimental list, we manipulated within items and within participants whether the sentence was correct or not. Therefore, among the total 12 sentences, only 6 sentences required a change of the word by adding a pseudo-suffix as -ino to generate *tacchino*, or by deleting it as -etta to produce *salvia*. In this task, we considered as dependent variables accuracy and RT.

Finally, Task 4, assessing lexical comprehension, was collectively administered. We asked children to choose a picture matching a target word (i.e., “tavolozza” *palette* or “pinna” *fin*) that was orally named by the experimenter, among a set of pictures representing: (i) the target item (a palette for “tavolozza”); (ii) in half of the experimental sets, an item that could be morphologically related (e.g., “tavolo” table with respect to “tavolozza” palette; Grossmann and Rainer, 2004); in the remaining half, an item phonologically related to the target (e.g., “penna” pen with respect to “pinna” fin); (iii) an item that could be semantically related (e.g., “pennello” *brush* with respect to palette); (iv) an item that could be semantically unrelated (e.g., “occhiali” *glasses*). Each participant was provided with a booklet reporting four pictures for each experimental item and was told to mark the correct one. The 26 testing items (of which the first two served as practice trials and were thus excluded from the analysis) are reported in Appendix A.

Importantly, all words employed in the experimental tasks were drawn from classical studies of morphological masked priming conducted in Italian (e.g., Marelli et al., 2013; Amenta et al., 2015) or reading experiments run on Italian fourth and fifth graders with and without reading difficulties (Traficante et al., 2014).

## Standardized Tests

Reading speed, accuracy, and comprehension scores were obtained from the administration of the following Italian standardized tests: MT-2 reading tests (Prove MT-2 di lettura per la scuola elementare, Cornoldi and Colpo, 2011), which provide accuracy and speed measures for passage reading and accuracy scores for passage comprehension; test of word and non-word reading (Prova di lettura di parole e non parole, Zoccolotti et al., 2005), in which speed and accuracy scores were computed for 30 multisyllabic (i.e., made up of three or more syllables) words balanced for frequency of use and 30 (multisyllabic) non-words.

## Procedure and Design

The morphological awareness tasks (except for Task 4), as well as the reading tests (except for text reading comprehension), were administered in individual sessions. The procedure was as follows. As for Task 1, in four practice trials children were first

trained to recognize whether the two words presented were related to each other. In the practice trials, after the simultaneous oral and visual presentation of the two words (e.g., “torta” *cake*, “tortina” *little cake*; or “colla” *glue*, “collina” *hill*), the female voice on the computer explained why two words were related to each other (i.e., “a little cake is a cake”) or not (“a hill is not a little glue”). In the experimental phase, the recorded voice asked after each pair of words whether the child thought they were related or not. Children were told to press as soon as possible a button on the keyboard marked with “Sì” *yes* or “No” *no*, to provide their answer.

Regarding Task 2, children received both oral and written instructions. They were told that they would see a picture and hear a voice naming it, then a picture representing an item semantically and morphologically related to the previous picture would appear on the screen. Their task was to name it. Again, there were six practice trials to make sure that children understood the task. For instance, after seeing the picture of a pizza and hearing a voice pronouncing it, children saw the image of a pizza restaurant. If they said “ristorante” *restaurant*, they were corrected and invited to describe it using a word “related to pizza,” namely “Pizzeria.”

As for Task 3, children were simply told to listen to a series of sentences describing the pictures appearing on the screen. At the end of each sentence they had to press a button if they detected an anomaly (“a mistake”) in the sentence, and if so, they had to correct it out loud. The experimenter recorded their answers and coded them off-line.

For production Tasks 2 and 3, non-target responses were coded as morphologically relevant or irrelevant depending on whether they involved a totally unrelated word from a morphological perspective (e.g., as in the case of “ristorante” instead of “pizzeria”), or whether the error referred to the choice of an incorrect suffix to derive the new word. For example, after hearing “sacco” *bag*, a child produced “\*sacchino” instead of “sacchetto” (*little bag*). Literally, “\*sacchino” is a morphologically well-formed word, but it does not exist in Italian.

To assess reading ability, participants had to read the lists of words and non-words and the passage according to their grade level. In the reading comprehension test, the participant had to silently read a text and answer multiple-choice questions, with the possibility of accessing the text. Speed (number of syllables read divided by time in seconds to read them) and accuracy (number of errors) were calculated. Raw accuracy scores were converted to standardized scores (z-scores).

In each testing session, administration of reading tests was interspersed with that of morphological awareness tasks. Therefore, participants were individually tested in two sessions, lasting approximately 20 min each. Only Task 4 (lexical comprehension task) and the standardized test of text comprehension were collectively administered to the whole group.

## RESULTS

### Data Treatment and Statistical Analysis

First RT data of morphological awareness tasks were trimmed to remove outliers. We excluded two types of outlier trials: outliers defined as any RTs shorter than 100 ms and outliers defined as

RTs that were 2.5 SDs slower than the relevant mean RT (Baayen and Milin, 2010). After excluding outliers, we calculated mean RTs across subjects. We could not add as an additional dependent variable “Non target responses” to Tasks 2 and 3, as we did not have enough data points to run the analysis. All statistical analyses were performed using the log-transformed data.

Both RT and accuracy data were fitted to a series of general linear models and mixed-effects models using the statistical environment R (R Core Team, 2014), and in particular the packages Rcmdr (Fox, 2017), lme4, and lmerTest (Kuznetsova et al., 2017). In each analysis, we tested whether a reading variable was significantly predicted by the performance (accuracy and speed) to morphological awareness tasks and to the lexical task. In this section, we report the results of the *t*-tests and a summary of the fixed effects of the final (linear and mixed effects) models. In each model, we used a stepwise model selection procedure to estimate whether the inclusion of the morphological variables considered (Task 1 accuracy and speed; Task 2 accuracy; Task 3 accuracy and speed; Task 4 accuracy) added information to the models' fit and had to be included. For each dependent variable, we started with a base model, and then added each individual factor. If adding each factor did not result in a significant gain of the model fit, we removed it from the final model (Baayen et al., 2008). In mixed-effects models (that involve a specification of the random effects structure too), we started with a base model that included only a by-participants and a by-items random intercept. Then, we tested whether the inclusion of a by-participants or by-items random slope for each significant factor improved the fit of the model in comparison with the base model. All the best fitting models

involved a basic random structure (i.e., a by-participants and by-items random intercept). For completeness sake, all the final models are available at this link ([https://docs.google.com/document/d/1NMX0A1gSkoy\\_VzlSH1AFClIGdgipvXGqr7S4C-3T9W0/edit?usp=sharing](https://docs.google.com/document/d/1NMX0A1gSkoy_VzlSH1AFClIGdgipvXGqr7S4C-3T9W0/edit?usp=sharing)).

## Analysis: L1 Competent vs. L1 Beginning Readers

Descriptive statistics (mean and SDs) for all the variables included in the study are reported in **Table 1** (morphological awareness tasks and lexical task) and in **Table 2** for standardized reading tests. For the sake of simplicity, we provide a short but hopefully clear summary of the statistical results we have obtained in the analysis that compares competent vs. beginning readers on **Table 3** (first two rows).

First, we compared the performance of monolingual competent vs. beginning readers on the experimental tasks and standardized tests by means of a series of independent samples comparisons (*t*-tests). Regarding the morphological awareness tasks, there was a significant difference only in RTs of the production Task 3 [ $t(39) = -2.086, p < 0.04$ ], with competent readers being significantly faster than beginning readers. Regarding reading tests, beginning readers' performance was, as expected, significantly slower, but not less accurate, when compared with competent readers: differences were significant with respect to reading times of non-words [ $t(39) = 2.164, p < 0.037$ ], words [ $t(39) = 3.576, p < 0.001$ ], and passage reading [ $t(39) = 2.667, p < 0.011$ ].

**TABLE 1** | Descriptive statistics (mean and SDs) of all variables for the morphological awareness tasks and for the lexical comprehension task in the L1 competent and beginning readers groups.

| Variable               | Comprehension Task 1                         |      |          |          | Production Task 2      |      | Production Task 3      |      |          |          | Lexical comprehension Task 4 |      |            |      |
|------------------------|----------------------------------------------|------|----------|----------|------------------------|------|------------------------|------|----------|----------|------------------------------|------|------------|------|
|                        | Accuracy (proportions of accurate responses) |      | RT       |          | Accuracy (proportions) |      | Accuracy (proportions) |      | RT       |          | z-Scores                     |      | Raw scores |      |
|                        | Mean                                         | SD   | Mean     | SD       | Mean                   | SD   | Mean                   | SD   | Mean     | SD       | Mean                         | SD   | Mean       | SD   |
|                        |                                              |      |          |          |                        |      |                        |      |          |          |                              |      |            |      |
| Competent readers (L1) | 0.80                                         | 0.15 | 4,290.89 | 1,124.40 | 0.75                   | 0.12 | 0.82                   | 0.13 | 7,542.94 | 2,158.00 | 0.35                         | 0.58 | 23.15      | 1.72 |
| Beginning readers (L1) | 0.75                                         | 0.14 | 4,255.97 | 507.83   | 0.69                   | 0.17 | 0.78                   | 0.10 | 9,214.06 | 2,495.42 | 0.27                         | 0.69 | 22.91      | 2.04 |

**TABLE 2** | Reading and spelling performances of L1 competent and beginning readers on the standardized reading tests.

| Variables              | Word   |      |            |      | Non-word |      |            |      | Passage reading |      |            |      | Text comprehension |      |
|------------------------|--------|------|------------|------|----------|------|------------|------|-----------------|------|------------|------|--------------------|------|
|                        | Speed  |      | Accuracy   |      | Speed    |      | Accuracy   |      | Speed           |      | Accuracy   |      | Accuracy           |      |
|                        | syll/s |      | (z-Scores) |      | syll/s   |      | (z-Scores) |      | syll/s          |      | (z-Scores) |      | (z-Scores)         |      |
|                        | Mean   | SD   | Mean       | SD   | Mean     | SD   | Mean       | SD   | Mean            | SD   | Mean       | SD   | Mean               | SD   |
| Competent readers (L1) | 3.33   | 1.37 | -0.07      | 1.40 | 1.59     | 0.59 | 0.37       | 0.81 | 2.90            | 1.18 | -0.51      | 0.78 | -0.29              | 0.88 |
| Beginning readers (L1) | 2.08   | 0.82 | -0.22      | 1.23 | 1.25     | 0.42 | 0.15       | 0.62 | 2.12            | 0.62 | -0.07      | 0.90 | -0.15              | 0.55 |

**TABLE 3** | Summary of the results of the GLMs that were conducted to test the predictive role of morphological awareness variables on reading skills.

|                      | Morphological awareness predictors |    |                   |                    |                    |                                 |
|----------------------|------------------------------------|----|-------------------|--------------------|--------------------|---------------------------------|
|                      | Comprehension Task 1               |    | Production Task 2 | Production Task 3  |                    | Lexical comprehension Task 4    |
|                      | Accuracy                           | RT | Accuracy          | Accuracy           | RT                 | z-Scores                        |
| L1 competent readers | –                                  | –  | –                 | Text comprehension | Text comprehension | Text comprehension word (speed) |
| L1 beginning readers | NW (speed)                         | –  | NW (speed)        | Word (accuracy)    |                    | Passage reading (accuracy)      |
|                      | NW (accuracy)                      |    | Word (speed)      |                    |                    | Text comprehension              |
|                      | Word (speed)                       |    |                   |                    |                    |                                 |
|                      | Word (accuracy)                    |    |                   |                    |                    |                                 |
| L2                   | Passage reading (speed)            | –  | NW (accuracy)     | NW (accuracy)      | –                  | Text comprehension              |

**TABLE 4** | Descriptive statistics (mean and SDs) of all variables for the morphological awareness tasks and for the lexical comprehension task in L2 children and in L1 peers matched for chronological age.

| Variable | Comprehension Task 1                         |      |          |          | Production Task 2      |      | Production Task 3      |      |          |          | Lexical comprehension Task 4 |      |            |      |
|----------|----------------------------------------------|------|----------|----------|------------------------|------|------------------------|------|----------|----------|------------------------------|------|------------|------|
|          | Accuracy (proportions of accurate responses) |      | RT       |          | Accuracy (proportions) |      | Accuracy (proportions) |      | RT       |          | z-Scores                     |      | Raw scores |      |
|          | Mean                                         | SD   | Mean     | SD       | Mean                   | SD   | Mean                   | SD   | Mean     | SD       | Mean                         | SD   | Mean       | SD   |
| L1       | 0.73                                         | 0.15 | 4,263.07 | 598.13   | 0.64                   | 0.14 | 0.76                   | 0.13 | 9,324.59 | 2,807.03 | 0.41                         | 0.46 | 23.17      | 1.34 |
| L2       | 0.63                                         | 0.14 | 4,697.88 | 1,232.86 | 0.48                   | 0.26 | 0.63                   | 0.13 | 8,092.66 | 1,355.49 | –1.05                        | 1.30 | 19.00      | 3.86 |

Second, a series of linear models were conducted to determine the specific contribution of morphological awareness measures on reading ability. In the models reported in this section of the paper and in all the subsequent ones, the sign of the coefficient assumes a positive value, denoting that the odds for an accurate/fast reading performance become larger when responses to the morphological tasks are more accurate/faster.

In the beginning readers group, the two reading fluency components (accuracy and speed) were differentially affected by morphological awareness. Regarding non-words, speed was significantly predicted by accuracy to the comprehension Task 1, e.g., establishing whether a pair of words was morphologically related or not (estimate = 0.892, SE = 0.698,  $t = 3.725$ ,  $p < 0.001$ ) and the production Task 2 (estimate = 1.668, SE = 0.604,  $t = 2.760$ ,  $p < 0.013$ ). The only significant predictor of non-words reading accuracy was again accuracy to the comprehension Task 1 (estimate = 0.635, SE = 0.766,  $t = 3.579$ ,  $p < 0.002$ ).

Word reading speed was significantly predicted by accuracy to the comprehension Task 1 (estimate = 0.890, SE = 1.347,  $t = 3.749$ ,  $p < 0.001$ ) and to the production Task 2, e.g., transforming a base word into a derived one (estimate = 0.678, SE = 1.166,  $t = 2.858$ ,  $p < 0.010$ ). Word reading accuracy was predicted by accuracy to the comprehension Task 1 (estimate = 4.462, SE = 1.346,  $t = 3.316$ ,  $p < 0.004$ ) and to the production Task 3 (e.g., correcting a short sentence by generating a new word that involved the same pseudo-stem but a different pseudo-suffix) (estimate = 4.921, SE = 1.871,  $t = 2.630$ ,  $p < 0.017$ ). Finally, accuracy to the lexical comprehension Task 4 appeared to be the unique significant predictor of the accuracy to passage reading (estimate = 0.713, SE = 0.190,  $t = 3.729$ ,  $p < 0.001$ ) and comprehension (estimate = 0.127, SE = 0.054,  $t = 2.359$ ,  $p < 0.0292$ ).

In the competent readers group, a general result indicated that reading fluency variables were not significantly predicted by performance to morphological awareness tasks, while reading comprehension was strongly influenced by accuracy and RT measures of the production Task 3 and by accuracy to the lexical Task 4. Regarding reading skills, only accuracy to the lexical Task 4 predicted word reading speed (estimate = 0.401, SE = 0.161,  $t = 2.495$ ,  $p < 0.022$ ).

Remarkably, reading comprehension scores were significantly predicted by both accuracy (estimate = 8.686, SE = 2.51,  $t = 3.457$ ,  $p < 0.003$ ) and RT data (estimate = 0.00029, SE = 0.001,  $t = 2.68$ ,  $p < 0.0164$ ) in the production Task 3, as well as by accuracy to the lexical comprehension Task 4 (estimate = 0.39, SE = 0.183,  $t = 2.123$ ,  $p < 0.049$ ). No other effect was found in this group.

To sum up, while reading speed and accuracy were positively related in first and second grade to morphological awareness tasks, in competent readers reading comprehension was predicted only by accuracy and speed measures in the production Task 3. In competent readers, the stronger predictor of both word reading and text comprehension appeared to be lexical competence (i.e., accuracy to the lexical Task 4). Overall current results suggest that in beginning readers reading fluency at word-level was significantly affected by morphological awareness (accuracy), while in competent ones accuracy and speed in a morphological task, together with accuracy to the lexical task, showed a predictive role for reading comprehension.

## Analysis: L1 vs. L2

The second research question concerned the relationship between morphological awareness and reading measures of L2 children in comparison with their monolingual peers. **Tables 4** and **5** report the mean and the SDs of the L2 and L1 groups on the

**TABLE 5** | Reading and spelling performances in L2 children and in L1 peers matched for chronological age on the standardized reading tests.

| Variables | Word  |      |          |      | Non-word |      |          |      | Passage reading |      |          |      | Text comprehension |      |
|-----------|-------|------|----------|------|----------|------|----------|------|-----------------|------|----------|------|--------------------|------|
|           | Speed |      | Accuracy |      | Speed    |      | Accuracy |      | Speed           |      | Accuracy |      | Accuracy           |      |
|           | Mean  | SD   | Mean     | SD   | Mean     | SD   | Mean     | SD   | Mean            | SD   | Mean     | SD   | Mean               | SD   |
| L1        | 2.75  | 1.20 | -0.35    | 1.13 | 1.42     | 0.44 | -0.44    | 0.70 | 2.66            | 0.86 | -0.76    | 0.86 | -0.35              | 0.90 |
| L2        | 2.58  | 1.49 | -1.61    | 1.82 | 1.59     | 0.61 | 0.43     | 1.12 | 2.05            | 0.90 | -3.95    | 4.53 | -0.46              | 0.33 |

morphological awareness measures and standardized tests of reading ability. **Table 3** (third and fourth rows) report a synthetic summary of the statistical results we have obtained when comparing L1 vs. L2 groups. Note that, due to the reduced number of L2 children (12), we did not conduct a separate analysis comparing beginning and competent readers in this sample. In addition, based on the evidence that L2 children might possibly show delayed achievements in reading when compared with L1 (cf. Murineddu et al., 2006), one might not safely assume that L2 children, in comparison with L1 ones, master decoding skills by third grade.

First, we tested whether the two groups differed on morphological awareness skills (accuracy and RT) as well as on reading ability by means of a series of independent mean comparisons (*t*-tests). As for morphological tasks, when we considered accuracy as dependent variable, L1 children performed better than L2 in basically all tasks [production Task 2:  $t(22) = 2.024$ ,  $p < 0.054$ ; production Task 3:  $t(22) = 2.384$ ,  $p < 0.026$ ; lexical comprehension Task 4:  $t(22) = 3.532$ ,  $p < 0.002$ ], except for the comprehension Task 1 (e.g., establishing whether a pair of words was morphologically related or not), where the difference did not reach significance [ $t(22) = 1.666$ ,  $p = 0.11$ ]. Interestingly, group difference was no longer significant when we considered the RT data of the morphological awareness tasks as a dependent variable. Regarding reading tests, L2 were significantly less accurate when compared with their L1 peers in word reading accuracy [ $t(22) = 2.037$ ,  $p < 0.054$ ]. However, interestingly, in non-word reading accuracy, L2 significantly outperformed L1 children [ $t(22) = -2.272$ ,  $p < 0.033$ ].

As the L2 group was not balanced for age, we further conducted a series of linear models and linear mixed-effects models to test whether the contribution of age (in months) and/or length of exposure to Italian (in months) could possibly interact and eventually overcome the effect of group. Therefore, by means of a models comparison procedure, we evaluated the contribution of all these factors (group, age, length of exposure to Italian) on morphological tasks as well as on reading ability. Regarding the morphological tasks, accuracy appeared to be significantly affected by length of exposure in Tasks 1–3; however, the first level effect of group remained significant (marginally in Task 1) and had to be kept in the models [Task 1: reference level: L2; estimate =  $-0.46$ , SE = 0.25,  $t = -1.83$ ,  $p < 0.06$ ; Task 2: reference level: L2; estimate =  $-1.121$ , SE = 0.322,  $t = -3.47$ ,  $p < 0.001$ ; Task 3: estimate =  $-1.011$ , SE = 0.23,  $t = -4.39$ ,  $p < 0.001$ ] as demonstrated by a series of mixed-effects models.

In the lexical comprehension Task 4, there was a significant effect of group [reference level: L2; estimate =  $-4.166$ , SE = 1.18,  $t = -3.532$ ,  $p < 0.001$ ], with chronological age and language exposure not contributing information to the model (all  $p$ 's  $> 0.45$ ). Similarly, the group differences found in the reading tests were confirmed even when we included length of exposure to the models [non-word reading accuracy: reference level: L2; estimate = 0.834, SE = 0.34,  $t = 2.46$ ,  $p < 0.022$ ; word reading accuracy: estimate =  $-1.229$ , SE = 0.0595,  $t = -2.064$ ,  $p < 0.0516$ ].

## Analysis: L2

Again, we conducted a series of general linear models to test whether measures of morphological awareness predicted reading skills in bilingual development too. Importantly, in all the models we controlled whether the length of exposure to Italian contributed or not to the models' fit or not. For simplicity sake, we will report only significant results related to the morphological awareness measures and reading tests. Regarding non-words, accuracy was significantly predicted by accuracy in the production Task 2 (estimate = 5.036, SE = 1.470,  $t = 3.427$ ,  $p < 0.007$ ) and partially by accuracy in Task 3 (estimate = 6.195, SE = 2.978,  $t = 2.080$ ,  $p < 0.067$ ). When we considered passage reading speed as a dependent variable, accuracy to the comprehension Task 1 significantly contributed to the model (estimate = 4.174, SE = 1.584,  $t = 2.634$ ,  $p < 0.025$ ). In addition, accuracy to the lexical comprehension Task 4 appeared to be the only significant predictor of the accuracy in the passage comprehension test (estimate = 0.573, SE = 0.019,  $t = 2.923$ ,  $p < 0.015$ ). No other significant effect was found.

Overall, in the L2 group, we observed a pattern of results that resembled to a certain extent that of L1 beginning readers: morphological awareness predicted non-word reading accuracy, as well as passage reading speed, indicating that it contributed to reading ability at word level. However, none of the morphological variables, except for lexical ability, predicted reading comprehension, thus suggesting that at least in this sample, morphological awareness might contribute to text-level decoding, but not to comprehension.

## DISCUSSION

In this study, we explored the extent to which morphological awareness affects reading fluency and comprehension in monolingual (L1 Italian) and bilingual (Arabic-Italian) children coming from low SES background. We experimentally tested this research

question by designing three tasks of morphological awareness: one of comprehension and two of production. In addition, we evaluated students on a lexical task and on a range of standardized reading fluency and comprehension tests. Our results provided evidence supporting the existence of a general correlation between an *explicit* measure of morphological awareness to word derivation and reading ability (Kirby et al., 2012). Remarkably, as far as we are aware this is the first demonstration in Italian since the bulk of the current studies was conducted so far in English or Dutch (languages with opaque orthography).

Our findings suggested that morphological awareness is strictly intertwined with reading ability, though this relationship appeared to evolve significantly along with age, with crucial variations across different developmental populations. Let us start by discussing the outcomes observed in the monolingual sample. The data about L1 children highlighted two main findings: morphological awareness seems to influence word recognition and decoding early on during reading development, while in the last grades of primary school it showed a higher predictive impact on comprehension processes.

The pattern of results in the L1 beginning readers approached previous studies suggesting that decomposition of complex words (or even multi-morphemic non-words) into morphemic units supports reading ability in younger readers (Burani et al., 2008; Marcolini et al., 2011; Traficante et al., 2011). According to the above-mentioned literature, morpheme-based reading might allow children to read units smaller than the whole word, but bigger than the grapheme or the syllable (see, for instance, Angelelli et al., 2014). Our study adds to the previous research, by providing compelling evidence about the fact that an *explicit* measure of morphological awareness could be accounted as a significant predictor of accuracy and speed in words, non-words, and passage reading. Therefore, this might be an evidence that poor morphological representations and decomposition skills in children who are learning to read might be causally related to problems in reading fluency.

Data of the competent readers revealed a remarkably different pattern of results: morphological awareness and mostly lexical ability (i.e., accuracy to the lexical Task 4) played a significant role in predicting reading comprehension skills. Conversely, there was no contribution of morphological awareness to reading fluency. That is, results based on skilled readers revealed that, once lexical access is automatized in reading, involving direct access to a lexical unit fully defined from an orthographic, morphological and semantic perspective, morphological knowledge as well as vocabulary support comprehension skills, presumably allowing readers to make inferences about complex words in the text.

Note that this pattern of findings might appear hard to reconcile with a well-accepted view under which the relation between morphological awareness and reading achievement increases with age and grade level (Nagy and Anderson, 1984; Anglin, 1993). By contrast, we observed that awareness of derivational morphology predicts reading fluency in first and second graders, while from third grade on, it appears to support only reading comprehension.

Even though we have observed that from third to fifth grade, morphological awareness no longer contributes to word

decoding, it is important to note that, according to the literature, some changes in the (implicit) processing of morphologically complex words might still occur in this time window. For example, Hasenäcker et al. (2017) indicate that morphemes progressively emerge as units of word recognition in the course of reading development in German children from 7 years of age up to 9, with peculiar differences between types of affixes (compound, suffixes, etc.). Dawson et al. (2017) report qualitative differences in the way English-speaking 7- to 9-year-old process complex words when compared with young and older adolescents, suggesting that (implicit) morphological processing continue to develop up to early adolescence.

In general, it is presumable that morphological knowledge, as a metalinguistic skill, entails not only a lexical-semantic and syntactic component but also a phonological processing instance. All these sources of morphological information are differentially at stake in specific stages of reading development. In early reading achievement, for instance, awareness of words composition relies over and above the level of phoneme, contributing to decoding skills (e.g., Mann and Singson, 2003; Deacon and Kirby, 2004); later in the development of reading, awareness of the lexical-semantic decomposition of a complex word seems to support comprehension only indirectly through word reading (e.g., Deacon et al., 2014). Under this view, our data provide further evidence supporting the idea that morphological awareness increasingly supports reading achievement along the course of development, but that the nature of its role evolves over time.

The second question that this study addressed was whether morphological awareness in L2 children would be able to predict reading skills to the same extent as in L1. In general, L2 children showed a significantly poorer performance when compared with L1 children with respect to both morphological awareness ability and reading skills. At a more fine-grained level of analysis, considering the performance in single tasks, L2 children underperformed compared with monolingual children in tasks assessing morphological awareness and lexical ability. In particular, accuracy in the production Tasks 2 and 3 was significantly lower than in L1 children, while no difference was found for RTs. As for the accuracy in these tasks, there were a number of non-target responses that involved the choice of an incorrect suffix to derive the new word (i.e., producing “sacchino” instead of “sacchetto” little bag). Even though this type of error occurred in both groups, it was more likely in L2 children (though the number of data points was too small to perform a reliable analysis), indicating that the performance to the production tasks depended to a great extent on children's lexical knowledge. Perhaps most interestingly, it also indicated that children relied on morphological rules to produce a new word. Future research should focus on disentangling the causes of difficulties that emerged in our L2 sample, possibly due to a lack in lexical knowledge, from their derivational morphological skills.

Overall, our findings are in line with a number of previous studies (cf. Goodwin et al., 2011; Ramirez et al., 2011). For instance, a research based on a population of Spanish-English bilinguals coming from low SES background (Park et al., 2014) revealed that L2 children underperformed monolinguals both on accuracy and RTs in two tasks of morphological awareness

(morpheme blending and morpheme generation). To account for the differences found on morphological awareness in these groups, the authors propose a bilingual lexical interference explanation. Namely, bilinguals activate two competing lexical entries for the same word (Grosjean, 2001; Marian and Spivey, 2003), which might additionally involve multiple derived representations in each language. However, building on the fact that results were based on bilinguals whose L2 lexical knowledge, as well as linguistic stimulation, appeared to be impoverished, it is possible that it was not bilingualism *per se* to hamper morphological processing. Recall indeed that there is strong evidence for a bilingual advantage on metalinguistic tasks, among which morphological awareness (Bialystok et al., 2014). Therefore, even in our study, it is possible that poorer morphological representation in bilingual children could be due to a reduced L2 vocabulary size, in particular with respect to those words used in our tasks that were somewhat less common in the everyday life.

Regarding standardized tests, L2 children's accuracy on word reading was significantly lower when compared with their monolingual peers. Conversely, in non-word reading, the significant difference indicated a better performance of L2 children. This pattern of results confirms that reading in bilingual children might be characterized by an over-reliance on sub-lexical processing mechanisms; this tendency seems to significantly facilitate non-word reading, while critically hindering word reading efficacy. Note that such finding is in line with previous research based on early L2 Italian pre-school children, which showed a performance on non-word repetition comparable to their monolingual peers. By contrast, their ability in other tasks of morpho-syntactic knowledge appeared significantly lower with respect to monolinguals, though not directly comparable to that of SLI children (Vender et al., 2016).

Again, note that it was not the aim of this article to disentangle whether word decomposition in reading was based on grapheme-phoneme correspondence or on the morphological structure. Indeed, in contrast to previous studies that manipulated on purpose the morphological structure of lexical reading stimuli, we relied on already existing standardized reading tests to evaluate children's reading ability. Our results simply indicate that morphological awareness stands out as a strong linguistic underpinning of lexical decoding in learning to read during the early years, or in older readers such as bilinguals coming from low SES, who still struggle to read.

With regard to the effects of morphological processing efficiency on bilingual reading, we observed that accuracy in morphological awareness tasks predicted non-word reading fluency, while efficiency in morphological awareness tasks was a significant predictor of passage reading accuracy. In addition, lexical ability (assessed by means of accuracy to the lexical Task 4) did not predict reading fluency, but only text comprehension skills. That is, lexical competence exerted its predictive effect only on text comprehension, suggesting that learners, whether bilingual or monolingual, nearly exclusively relied on vocabulary to comprehend a text. Such pattern of results suggests that, as in L1 beginning readers, in L2 children morphological processing supports word (text) and non-word decoding, while vocabulary skills seem to be involved in text comprehension. Therefore,

one might claim that reading at word-level appears to rely on decomposition processes that are supported by morphological skills, while comprehending a text is almost exclusively predicted by the lexical competence of the reader at least in L1 beginning readers and in L2 children.

Regarding lexical ability, in contrast to previous research revealing a strong correlation between L2 vocabulary size and the probability of a correct reading aloud performance (Primativo et al., 2013), we did not replicate such finding, as lexical ability appeared to be involved only in text comprehension also in L2 children. However, such difference could possibly be due to the fact that, in comparison with previous studies that contrasted both a receptive and expressive component of vocabulary, in the current one we tested only lexical comprehension.

In general, given the different role that morphological awareness appears to play along the course of reading development, it should be important to consider such skill to improve the ability to read not only in children who are learning but also in populations who struggle with reading, such as bilinguals whose L2 knowledge is somewhat impoverished. In the literature, a range of possible interventions based on morphological awareness instructions is reported. Kuo and Anderson (2006), for instance, suggest that, in reading, placing a syllable break based on the morphological structure of the word would help children to recognize the deep structure of a word. To exemplify, the pronunciation of *-ive* in *suggestive* would involve a syllable break as it is a derivational suffix but it would not if it is part of a word such as *arriv-e*. Again, for words like *peeled* (involving two morphemes: peel-ed) and *field* (one morpheme), which sound similar but involve a different spelling, stressing the -ed morpheme in *peeled* would offer students a clarification of the different spellings (Nunes et al., 2006). By doing so, children will be more conscious of the fact that apparently similar sub-units (or pseudo-units) of words involve a different and specific relationship to grammar. Another evidence suggesting the benefits coming from hyphenation in marking morphological structure come from an eye-tracking study by Häikiö et al. (2011). These authors showed that Finnish children in the early stage of reading processed more easily hyphenated compound words than concatenated ones, suggesting that they strongly rely on morphemes when processing compounds (see Colé et al., 2012 for similar evidence from French children). In general, knowledge of the deep morphological structure of a word would allow children to improve their phonological processing difficulties in reading and spelling (Goodwin and Ahn, 2010).

We are conscious of the fact that this research presents some limitations. First, we cannot exclude that the morphological tasks used in this study, involving real words' root and real derivational suffixes were meant to measure morphological awareness could, in fact, be influenced by lexical knowledge of the suffixed and/or pseudo-suffixed words. In our study, by using accuracy to the lexical Task 4 as a measure of vocabulary skills, we were able to disentangle any influence of lexical competence on reading fluency that was exerted by the morphological variables. Note however that, since we used a non-standardized measure of lexical ability, that was not yet correlated with other (standardized) measures of vocabulary, we might not be completely sure whether such lexical measure assessed vocabulary size or a

more general lexical comprehension ability. As a consequence, one might not exclude that the difference that emerged between L1 and L2 groups could have been due to lexical knowledge (or vocabulary size) and not only to morphological knowledge *per se*. Note, however, that this might be considered a problem for most of the tasks in the literature on (derivational and inflectional) morphological awareness (cf. Carlisle, 2000; Singson et al., 2000; Kieffer and Lesaux, 2012), involving the transformation of real roots into new derived words by using real suffixes. Therefore, even in those studies, one cannot exclude the contribution of lexical competence on morphological awareness achievement.

One way out of this dilemma would be creating a series of morphological awareness tasks involving pseudoword material, attesting that children are able to apply productive rules to generate derivations of novel words. If the child can provide the correct derived form of a non-existing word, we might have rather uncontroversial evidence that the child possesses and is able to apply the derivational rules. To date, morphological tasks involving non-words appear to be mostly used to test competence for inflectional rules (as in the popular Wug Test; see in Italian, for instance, Vender et al., 2017). It appears therefore that a morphological awareness task constructed on nonsense material might represent a promising future path of our work.

Second, we did not directly test children on measures of linguistic processing in L1 and L2, such as phonological skills and awareness, syntactic competence and working memory, as well as a standardized measure of vocabulary. By evaluating linguistic abilities, one could possibly assess to what extent other linguistic components interact with morphological awareness in reading achievement along the course of development. In particular, it would be of interest to test the relationship between morphological processing and phonological elaboration. As we proposed previously, it is possible that in the initial stages of learning to read the awareness of the linguistic units of a word is strictly intertwined with its phonemic representation. Therefore, the relationship of phonological and morphological awareness deserves further investigation in future studies, also with reference to the possible interventions targeting the extent to which morphological processing might improve depleted phonological abilities.

Considering the bilingual group, to draw clear conclusions regarding differences between them and their monolingual peers, a more detailed account of the role of proficiency in their L1 as

well as L2 should be more specifically addressed in further studies. In addition, this study does not offer an exhaustive evaluation of the role of SES and of cumulative length of exposure to Italian (rather than traditional; cf. Unsworth, 2013). To reconcile with these limitations, note, however, that all the children were selected from the same school and therefore they lived in the same area and had comparable educational exposure. As reported earlier, an important contribution in this direction would be disentangling to what extent (L2) children show the ability to apply derivational rules irrespective of their lexical knowledge. In this study, we observed such tendency, however, due to the low number of errors produced we could not analyze the data.

In summary, this findings suggest that morphological awareness is a crucial construct to consider in reading development not only in monolingual children but in bilingual too, as it might offer an independent contribution earlier to reading decoding and comprehension in later grades of primary school.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of The Ethical committee of the University of Milano-Bicocca with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethical committee of the University of Milano-Bicocca.

## AUTHOR CONTRIBUTIONS

MV: conception and design of the work; data collection; data analysis and interpretation; drafting the article. EP: critical revision of the article. MV and EP: final approval of the version to be publishable.

## ACKNOWLEDGMENTS

This publication has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 641858 MSCA ITN PREDICTABLE. Finally, we would like to thank Federica Francesca Ioppolo for her assistance with data collection. We also acknowledge all the children, teachers, and administrators who participated in the study or facilitated the research.

## REFERENCES

- Amenta, S., Marelli, M., and Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: early morpho-orthographic and morpho-semantic effects in sentence reading. *J. Exp. Psychol.* 41, 1587–1596. doi:10.1037/xlm0000104
- Angelelli, P., Marinelli, C. V., and Burani, C. (2014). The effect of morphology on spelling and reading accuracy: a study on Italian children. *Front. Psychol.* 5:1373. doi:10.3389/fpsyg.2014.01373
- Angelelli, P., Marinelli, C. V., De Salvatore, M., and Burani, C. (2017). Morpheme-based reading and spelling in Italian children with developmental dyslexia and dysorthography. *Dyslexia* 23, 387–405. doi:10.1002/dys.1554
- Angelelli, P., Notarnicola, A., Judica, A., Zoccolotti, P., and Luzzatti, C. (2010). Spelling impairments in Italian dyslexic children: phenomenological changes in primary school. *Cortex* 46, 1299–1311. doi:10.1016/j.cortex.2010.06.015

- Anglin, J. M. (1993). "Vocabulary development: a morphological analysis," in *Monograph of the Society for Research in Child Development* (Cambridge, UK: Cambridge University Press).
- Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual route model. *J. Mem. Lang.* 37, 94–117. doi:10.1006/jmla.1997.2509
- 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
- Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28. doi:10.21500/20112084.807
- Baddeley, A. D., Gathercole, S. E., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi:10.1037/0033-295X.105.1.158

- Bellocchi, S., and Genesee, F. (2012). L'apprendimento della lettura in bambini scolarizzati in una seconda lingua: traiettorie evolutive tipiche e difficoltà. *Psicologia Clinica dello Sviluppo* 3, 481–506. doi:10.1449/38836
- Berko, J. (1958). The child learning of English morphology. *Word* 14, 150–177. doi:10.1080/00437956.1958.11659661
- Bialystok, E., Craik, F. I. M., and Luk, G. (2008). Lexical access in bilinguals: effects of vocabulary size and executive control. *J. Neurolinguistics* 21, 522–538. doi:10.1016/j.jneuroling.2007.07.001
- Bialystok, E., Peets, K. F., and Moreno, S. (2014). Producing bilinguals through immersion education: development of metalinguistic awareness. *Appl. Psycholinguist.* 35, 177–191. doi:10.1017/S0142716412000288
- Burani, C. (2010). Word morphology enhances reading fluency in children with developmental dyslexia. *Lingue e Linguaggio* 9, 177–198. doi:10.1418/33326
- Burani, C., Marcolini, S., De Luca, M., and Zoccolotti, P. (2008). Morpheme-based reading aloud: evidence from dyslexic and skilled Italian readers. *Cognition* 108, 243–262. doi:10.1016/j.cognition.2007.12.010
- Carlisle, J. F. (1995). “Morphological awareness and early reading achievement,” in *Morphological Aspects of Language Processing*, ed. L. Feldman (Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.), 189–209.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: impact on reading. *Read. Writ.* 12, 169–190. doi:10.1023/A:1008131926604
- Carlisle, J. F. (2007). “Fostering morphological processing, vocabulary development, and reading comprehension,” in *Vocabulary Acquisition: Implications for Reading Comprehension*, eds R. K. Wagner, A. E. Muse, and K. R. Tannenbaum (New York: Guilford Press), 78–103.
- Carlisle, J. F., and Katz, L. (2006). Effects of word and morpheme familiarity on reading of derived words. *Read. Writ.* 19, 669–693. doi:10.1007/s11145-005-5766-2
- Casalis, S., Cole, P., and Sopo, D. (2004). Morphological awareness in developmental dyslexia. *Ann. Dyslexia* 54, 114–138. doi:10.1007/s11881-004-0006-z
- Casalis, S., and Louis-Alexandre, M. F. (2000). Morphological analysis, phonological analysis and learning to read French: a longitudinal study. *Read. Writ.* 12, 303–335. doi:10.1023/A:1008177205648
- Colé, P., Bouton, S., Leuwers, C., Casalis, S., and Sprenger-Charolles, L. (2012). Stem and derivational-suffix processing during reading by French second and third graders. *Appl. Psycholinguist.* 33, 97–120. doi:10.1017/S0142716411000282
- Cornoldi, C., and Colpo, G. (2011). *Prove di Lettura MT-2 per la Scuola Primaria (en. tr. Reading Test for Primary School)*. Firenze: Giunti O.S.
- Crepaldi, D., Rastle, K., Coltheart, M., and Nickels, L. (2010). ‘Fell’ primes ‘fall’, but does ‘bell’ prime ‘ball’? Masked priming with irregularly-inflected primes. *J. Mem. Lang.* 63, 83–99. doi:10.1016/j.jml.2010.03.002
- Dawson, N., Rastle, K., and Ricketts, J. (2017). Morphological effects in visual word recognition: children, adolescents, and adults. *J. Exp. Psychol. Learn. Mem. Cogn.* doi:10.1037/xlm0000485
- Deacon, S. H., Kieffer, M. J., and Laroche, A. (2014). The relation between morphological awareness and reading comprehension: evidence from mediation and longitudinal models. *Sci. Stud. Read.* 18, 432–451. doi:10.1080/10888438.2014.926907
- Deacon, S. H., and Kirby, J. R. (2004). Morphological awareness: just “more phonological”? The roles of morphological and phonological awareness in reading development. *Appl. Psycholinguist.* 25, 223–238. doi:10.1017/S0142716404001110
- Derwing, B. L., Smith, M. L., and Wiebe, G. E. (1995). “On the role of spelling in morpheme recognition: experimental studies with children and adults,” in *Morphological Aspects of Language Processing*, ed. L. B. Feldman (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 3–27.
- Fowler, A. E. (1991). “How early phonological development might set the stage for phoneme awareness,” in *Phonological Processes in Literacy: A Tribute to Isabelle Y. Liberman*, eds S. A. Brady and D. P. Shankweiler (Hillsdale, NJ: Lawrence Erlbaum Associates), 97–117.
- Fowler, A. E., and Liberman, I. Y. (1995). “The role of phonology and orthography in morphological awareness,” in *Morphological Aspects of Language Processing*, ed. L. Feldman (Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.), 157–188.
- Fox, J. (2017). *Using the R Commander: A Point-and-Click Interface for R*. Boca Raton FL: Chapman and Hall/CRC Press.
- Gathercole, S. E., and Baddeley, A. D. (1989). “The role of phonological memory in the development of complex verbal skills,” in *Brain and Reading*, ed. C. von Euler (London: Macmillan Press), 245–256.
- Gathercole, S. E., and Baddeley, A. D. (1993). Phonological working memory: a critical building block for reading development and vocabulary acquisition? *Eur. J. Psychol. Educ.* 8, 259–272. doi:10.1007/BF03174081
- Goodwin, A. P., and Ahn, S. (2010). A meta-analysis of morphological interventions: effects on literacy achievement of children with literacy difficulties. *Ann. Dyslexia* 60, 183–208. doi:10.1007/s11881-010-0041-x
- Goodwin, A. P., Huggins, A., Carlo, M., Malabonga, V., Kenyon, D., Louguit, M., et al. (2011). Development and validation of extract the base: an English derivational morphology test for third through fifth grade monolingual students and Spanish-speaking English language learners. *Lang. Test.* 29, 265–289. doi:10.1177/0265532211419827
- Grosjean, F. (2001). “The bilingual’s language modes,” in *One Mind, Two Languages: Bilingual Language Processing*, ed. J. Nicol (Oxford: Blackwell), 1–22.
- Grossmann, M., and Rainer, F. (2004). *La formazione delle parole in italiano*. Tübingen: Niemeyer.
- Häikiö, T., Bertram, R., and Hyönä, J. (2011). The development of whole-word representations in compound word processing: evidence from eye fixation patterns of elementary school children. *Appl. Psychol.* 32, 533–551. doi:10.1017/S0142716411000208
- Hasenäcker, J., Schröter, P., and Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *J. Exp. Psychol. Learn. Mem. Cogn.* 43, 1093–1108. doi:10.1037/xlm0000353
- Kieffer, M. J., Biancarosa, G., and Mancilla-Martinez, J. (2013). Roles of morphological awareness in English reading comprehension for Spanish-speaking language minority learners: exploring partial mediation by vocabulary and reading fluency. *Appl. Psycholinguist.* 34, 697–725. doi:10.1017/S0142716411000920
- Kieffer, M. J., and Lesaux, N. K. (2012). Knowledge of words, knowledge about words: dimensions of vocabulary in first and second language learners in sixth grade. *Read. Writ.* 25, 347–373. doi:10.1007/s11145-010-9272-9
- Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., and Parrila, R. (2012). Children’s morphological awareness and reading ability. *Read. Writ.* 25, 389–410. doi:10.1007/s11145-010-9276-5
- Kovelman, I., Baker, S. A., and Pettito, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism* 11, 203–223. doi:10.1017/S1366728908003386
- Kuo, L., and Anderson, R. C. (2006). Morphological awareness and learning to read: a cross-language perspective. *Educ. Psychol.* 41, 161–180. doi:10.1207/s15326985ep4103\_3
- Kuo, L. J., Ramirez, G., de Marin, S., Kim, T. J., and Unal-Gezer, M. (2017). Bilingualism and morphological awareness: a study with children from general education and Spanish-English dual language programs. *Educ. Psychol.* 37, 94–111. doi:10.1080/01443410.2015.1049586
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi:10.18637/jss.v082.i13
- Mann, V., and Singson, M. (2003). “Linking morphological knowledge to English decoding ability: large effects of little suffixes,” in *Reading Complex Words: Cross-Language Studies*, eds E. Assink and D. Santa (Dordrecht, The Netherlands: Kluwer), 1–25.
- Marcolini, S., Traficante, D., Zoccolotti, P., and Burani, C. (2011). Word frequency modulates morpheme-based reading in poor and skilled Italian readers. *Appl. Psycholinguist.* 32, 513–532. doi:10.1017/S0142716411000191
- Marelli, M., Amenta, S., Morone, E. A., and Crepaldi, D. (2013). Meaning is in the beholder’s eye: morpho-semantic effects in masked priming. *Psychon. Bull. Rev.* 20, 534–541. doi:10.3758/s13423-012-0363-2
- Marian, V., and Spivey, M. (2003). Comparing bilingual and monolingual processing of competing lexical items. *Appl. Psycholinguist.* 24, 173–193. doi:10.1017/S0142716403000092
- Marslen-Wilson, W., and Tyler, L. (1997). Dissociating types of mental computation. *Nature* 387, 592–594. doi:10.1038/42456
- Murineddu, M., Duca, V., and Cornoldi, C. (2006). Difficoltà di apprendimento scolastico degli studenti stranieri. *Difficoltà di apprendimento* 12, 49–70.
- Nagy, W., and Anderson, R. (1984). The number of words in printed school English. *Read. Res. Q.* 19, 304–330. doi:10.2307/747823
- Nagy, W., Berninger, V. W., and Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *J. Educ. Psychol.* 98, 134–147. doi:10.1037/0022-0663.98.1.134
- Nagy, W. M., and Townsend, D. (2012). Words as tools: ‘learning academic vocabulary’ as language acquisition. *Read. Res. Q.* 47, 91–108. doi:10.1002/RRQ.011

- Nunes, T., Bryant, P., Pretzlik, U., Burman, D., Bell, D., and Gardner, S. (2006). "An intervention program for classroom teaching about morphemes: Effects on the children's vocabulary," in *Improving Literacy by Teaching Morphemes*, eds T. Nunes and P. Bryant (London and New York: Routledge), 121–134.
- Orsolini, M., Fanari, R., Tosi, V., de Nigris, B., and Carrier, R. (2006). From phonological recoding to lexical reading: a longitudinal study on reading development in Italian. *Lang. Cogn. Process.* 21, 576–607. doi:10.1080/01690960500139355
- Park, J., Wiseheart, R., and Ritter, M. (2014). The roles of the speed and accuracy of morphological processing in the reading comprehension of Spanish-speaking language minority learners. *Int. J. Appl. Ling. Eng. Lit.* 3, 12–27. doi:10.7575/aiac.ijalel.v3n.1p.12
- Primativo, S., Rinaldi, P., O'Brien, S., Paizi, D., Arduino, L. S., and Burani, C. (2013). Bilingual vocabulary size and lexical reading in Italian. *Acta Psychol.* 144, 554–562. doi:10.1016/j.actpsy.2013.09.011
- Proctor, C. P., and Mo, E. (2009). The relationship between cognate awareness and English comprehension among Spanish–English bilingual fourth grade students. *TESOL Q.* 43, 126–136. doi:10.1002/j.1545-7249.2009.tb00232.x
- R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramirez, G., Chen-Bumgardner, B., Geva, E., and Luo, Y. (2011). Morphological awareness and vocabulary in ELLs. *Appl. Psycholinguist.* 32, 601–618. doi:10.1017/S0142716411000233
- 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
- Rispens, J. E., McBride-Chang, C., and Reitsma, P. (2008). Morphological awareness and early and advanced word recognition and spelling in Dutch. *Read. Writ.* 21, 587–607. doi:10.1007/s11145-007-9077-7
- Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L. L., and Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in grades 4, 6, and 8. *J. Exp. Child Psychol.* 102, 96–113. doi:10.1016/j.jecp.2008.01.004
- Saiegh-Haddad, E., and Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English-Arabic bilingual children. *Read. Writ.* 21, 481–504. doi:10.1007/s11145-007-9074-x
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools.
- Singson, M., Mahony, D., and Mann, V. (2000). The relation between reading ability and morphological skills: evidence from derivational suffixes. *Read. Writ.* 12, 219–252. doi:10.1023/A:1008196330239
- Tobia, V., and Marzocchi, G. M. (2014). Predictors of reading fluency in Italian orthography: evidence from a cross-sectional study of primary school students. *Child Neuropsychol.* 20, 449–469. doi:10.1080/09297049.2013.814768
- Tong, X., Deacon, S., Kirby, J. R., Cain, K., and Parrila, R. (2011). Morphological awareness: a key to understanding poor reading comprehension in English. *J. Educ. Psychol.* 103, 523–534. doi:10.1037/a0023495
- Traficante, D., Marcolini, S., Luci, A., Zoccolotti, P., and Burani, C. (2011). How do roots and suffixes influence reading of pseudowords: a study of young Italian readers with and without dyslexia. *Lang. Cogn. Process.* 26, 777–793. doi:10.1080/01690965.2010.496553
- Traficante, D., Marelli, M., Luzzatti, M., and Burani, C. (2014). Influence of verb and noun bases on reading aloud derived nouns: evidence from children with good and poor reading skills. *Read. Writ.* 27, 1303–1326. doi:10.1007/s11145-013-9488-6
- Tressoldi, P. E., Stella, G., and Faggella, M. (2001). The development of reading speed in Italians with dyslexia: a longitudinal study. *J. Learn. Disabil.* 34, 414–417. doi:10.1177/002221940103400503
- Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: the case of Dutch gender. *Bilingualism* 16, 86–110. doi:10.1017/S1366728912000284
- Vender, M., Garraffa, M., Sorace, A., and Guasti, M. T. (2016). How early L2 children perform on Italian clinical markers of SLI: a study of clitic production and nonword repetition. *Clin. Linguist. Phon.* 30, 150–169. doi:10.3109/02699206.2015.1120346
- Vender, M., Mantione, F., Savazzi, S., Delfitto, D., and Melloni, C. (2017). Inflectional morphology and dyslexia: Italian children's performance in a nonword pluralization task. *Ann. Dyslexia* 1–26. doi:10.1007/s11881-017-0152-8
- Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., and Martelli, M. (2009). Reading development in an orthographically regular language: effects of length, frequency, lexicality and global processing ability. *Read. Writ.* 22, 1053–1079. doi:10.1007/s11145-008-9144-8
- Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., and Spinelli, D. (2005). *Prova di lettura di parole e non parole*. Roma: IRCCS Fondazione Santa Lucia.

**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 © 2018 Vernice and Pagliarini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner 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.

## APPENDIX A

Task 1. List of the 32 experimental sets of words involving an opaque (1–16) vs. transparent morphological relationship (17–32). [English translation in brackets].

1. mulo (mule) mulino (mill); 2. botto (blow) bottone (button); 3. pulce (flea), pulcino (chick); 4. fiore (flower), fioretto (foil); 5. matto (mad), mattino (morning); 6. spunto (cue), spuntino (snack); 7. burro (butter), burrone (ravine); 8. bolla (bubble), bolletta (bill); 9. pollo (chicken), pollice (thumb); 10. latte (milk), lattuga (lettuce); 11. spina (thorn), spinacio (spinach); 12. lente (lens), lenticchia (lentil); 13. riva (shore), rivale (rival); 14. lava (lava), lavagna (blackboard); 15. tappo (cap), tappeto (carpet); 16. pista (trail), pistola (gun); 17. fieno (hay), fienile (barn); 18. calcio (soccer), calciatore (soccer player); 19. sasso (stone), sassata (throwing of a stone); 20. neve (snow) nevicata (snowfall); 21. tovaglia (cloth), tovagliolo (napkin); 22. forno (oven), fornaio (baker); 23. giardino (garden), giardinaggio (gardening); 24. cane (dog), canile (kennel); 25. lago (lake), laghetto (pond); 26. tubo (tube), tubetto (little tube); 27. zaino (backpack) zainetto (small backpack); 28. porta (door) portone (doorway); 29. anello (ring) anellino (little ring); 30. casa (house) casina (small house); 31 fontana (fountain) fontanella (small fountain); 32. asino (donkey) asinello (*little donkey*).

Task 2. List of the 16 sets of words (prime and target). [English translation in brackets].

Gelato (ice-cream), gelataio (ice-cream man); 2. Giardino (garden), giardiniere (gardener); 3 pane (bread), panettiere (baker); 4 dente (tooth), dentista (dentist); 5 libro (book), libreria (bookshop); 6 campana (bell) campanile (bell tower); 7 ghiaccio (ice) ghiacciolo (ice lolly); 8 gioco (game) giocattolo (toy); 9 borsa (bag) borsetta (handbag); 10 cesto (basket) cestino (trash can); 11 pentola (pot) pentolone (cauldron); 12 sacco (bag) sacchetto (small bag); 13 cappello (hat) cappellino (cap); 14 cioccolato (chocolate) cioccolatino (chocolate praline); 15 tazza (cup) tazzina (small cup); 16 villa (house) villetta (small house).

Task 3. List of the experimental materials used in the morphological awareness production task. [English translation in brackets].

Le piante amano il sole. I suoi raggi ne favoriscono la crescita/crescenza. [Plants love the sun. Its rays favor its growth/growth (unusual)].

In estate andiamo nel bosco a raccogliere i lamponi/lampi. [In summer we go on the countryside to pick up the raspberries/thunders].

Il caffè è in dispensa nel suo barattolo/baratto. [Coffee is in the pantry in its jar/barter].

Apreno tutto il rubinetto l'acqua esce con un bel getto/gettone. [By opening up the tap the water comes out with a nice jet/coin].

Il cane lo ha morso al polpaccio/polpo. [The dog has bitten him on the calf/octopus].

Carlo ha preparato l'arrosto di tacchino/tacco. [Carlo has prepared roasted turkey/heel].

La casa è sollevata dal tifone/tifo. [The house was raised by the typhoon/typhus].

Intorno allo stadio si sono verificati degli scontri/scontrini. [Around the stadium there have been clashes/sales receipts].

Nel suo orto Silvia coltiva la salvia/salvietta. [In his garden Silvia cultivates sage/towel].

In quella foto il nonno indossava una bombetta/bombola. [In that photo granpa wore a bowler hat/tank].

Il quadro è appoggiato sul cavalletto/cavallo. [The picture is resting on the easel/horse].

Sabato Marina è andata al circo/circuito. [Saturday Marina went to the circus/circuit].

Task 4. List of experimental materials used in the lexical comprehension task. [English translation in brackets]. In each row of the following list, the first word refers to the target item that was orally named by the experimenter; the second one refers to an item morphologically (items 1, 3, 4, 5, 7, 12, 15, 16, 18, 19, 21, 23, 26)/phonologically (items 2, 6, 8, 9, 10, 11, 13, 14, 17, 20, 22, 24, 25) related to the target; the third one to an item semantically related to the target; the fourth to an item unrelated to the target.

- (1) ombrello [umbrella], ombra [shadow], bastone [stick], albero [tree];
- (2) petali [petals], pedale [pedal], foglia [leaf], pentola [pan];
- (3) ortaggi [vegetables], orto [vegetable garden], frutta [fruits], vaso [flower pot];
- (4) bagno [bathroom], bagnino [lifeguard], acqua [water], guardia [policemen];
- (5) campanile [bell tower], campana [bell], chiesa [church], mongolfiera [hot-air balloon];
- (6) calice [chalice], camice [doctor's coat], tazza [cup], teiera [teapot];
- (7) tubo [pipe], tubetto (dentifricio) [tube], canale (d'acqua) [gutter], boccia (pesci) [fish bowl];
- (8) ardere [burn], radere [shave], fulminare [strike (by lightning)], nascere (pulcino) [to be born];
- (9) vitello [calf], vite [grapevine], mucca [cow], gallina [chicken];
- (10) stoppino [wick], stop [stop (road sign)], torcia [torch], regalo [present];
- (11) gomitolo [ball of wool], gomito [elbow], ago e filo [needle and thread], coltello [knife];
- (12) manubrio [handlebars], mano [hand], volante [steering wheel], gomito [elbow];
- (13) pinna [fin], penna [pen], balena [whale], latte [milk];
- (14) lampo [lightening], lampone [raspberry], pioggia [rain], cavallo [horse];
- (15) palo [pole], paletto [stake], tronco [trunk], cime montuose [mountain peaks];
- (16) coppetta [bowl], coppa [cup (winner)], tazza [mug], scacchi [chess];
- (17) colla [glue], collina [hill], forbici [scissors], albero [tree];
- (18) lancette [hands (clock)], lancia [spear], ago [needle], ascia [ax];
- (19) scuola [school], scolaro [pupil], libri [books], letto [bed];
- (20) boccia [fish bowl], doccia [shower], cuccia [dog's bed], chiesa [church];
- (21) pallottola [bullet], palla [ball], pistola [gun], torta [cake];
- (22) castagna [chestnut], castoro [beaver], ghianda [acorn], tazzina [mug];

- 
- (23) tavolozza [palette], tavolo [table], pennello [brush], occhiali [glasses];
- (24) nido [nest], dito [finger], pulcini [chicks], scoiattolo [squirrel];
- (25) vela [sail boat], tela [canvas], nave [boat], mongolfiera [hot-air balloon];
- (26) maglietta [t-shirt], maglione [sweater], gonna [skirt], mela [apple].



# Assessing the Formation of Experience-Based Gender Expectations in an Implicit Learning Scenario

Anton Öttl\* and Dawn M. Behne

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

## OPEN ACCESS

### Edited by:

Maria Garraffa,  
Heriot-Watt University,  
United Kingdom

### Reviewed by:

Sandy Caffarra,  
Basque Center on Cognition, Brain  
and Language, Spain  
LouAnn Gerken,  
University of Arizona, United States

### \*Correspondence:

Anton Öttl  
anton.oettl@ntnu.no

### Specialty section:

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

**Received:** 12 May 2017

**Accepted:** 17 August 2017

**Published:** 07 September 2017

### Citation:

Öttl A and Behne DM (2017)  
Assessing the Formation of  
Experience-Based Gender  
Expectations in an Implicit Learning  
Scenario. *Front. Psychol.* 8:1485.  
doi: 10.3389/fpsyg.2017.01485

The present study investigates the formation of new word-referent associations in an implicit learning scenario, using a gender-coded artificial language with spoken words and visual referents. Previous research has shown that when participants are explicitly instructed about the gender-coding system underlying an artificial lexicon, they monitor the frequency of exposure to male vs. female referents within this lexicon, and subsequently use this probabilistic information to predict the gender of an upcoming referent. In an explicit learning scenario, the auditory and visual gender cues are necessarily highlighted prior to acquisition, and the effects previously observed may therefore depend on participants' overt awareness of these cues. To assess whether the formation of experience-based expectations is dependent on explicit awareness of the underlying coding system, we present data from an experiment in which gender-coding was acquired implicitly, thereby reducing the likelihood that visual and auditory gender cues are used strategically during acquisition. Results show that even if the gender coding system was not perfectly mastered (as reflected in the number of gender coding errors), participants develop frequency based expectations comparable to those previously observed in an explicit learning scenario. In line with previous findings, participants are quicker at recognizing a referent whose gender is consistent with an induced expectation than one whose gender is inconsistent with an induced expectation. At the same time however, eyetracking data suggest that these expectations may surface earlier in an implicit learning scenario. These findings suggest that experience-based expectations are robust against manner of acquisition, and contribute to understanding why similar expectations observed in the activation of stereotypes during the processing of natural language stimuli are difficult or impossible to suppress.

**Keywords:** implicit learning, artificial language, frequencies of exposure, visual world eyetracking, gender representations, categorization, experience-based probabilities

## 1. INTRODUCTION

When processing a word referring to a human being, we typically activate an expectation as to whether it refers to a female or a male person. For most, the word “nurse” likely triggers a female representation, whereas “mechanic” likely triggers a male representation (see Misersky et al., 2013 for a survey of estimated gender distributions across several languages). While such

expectations can be traced to societal stereotypes and also to actual gender distributions (Gygax et al., 2016), certain languages additionally provide grammatical gender cues that may or may not be consistent with stereotypical information. For example, in Spanish grammatical gender is typically marked by the determiner “*el*” (masculine) or “*la*” (feminine), as well as by means of suffixation, as in “*camarero*” (male waiter) vs. “*camarera*” (female waiter). Though grammatical cues in principle could override stereotypical information, previous research has shown to the contrary that stereotypical gender information is activated automatically, even in cases where it is not needed for discourse coherence (Pyykkönen et al., 2010). One challenge to examining the interplay between linguistic and experience-based sources of gender information at the lexical level is the complexity of gender coding systems found in natural languages, as well as the stereotypes associated with them. The present study therefore employs an artificial language paradigm, in which aspects of interest, gender-coding and experience-based expectations, can be simulated in a laboratory setting and studied in isolation. While previous research has shown that an artificial language paradigm is adequate for studying the formation of new representations (Magnuson et al., 2003) and also more specifically the emergence of probabilistic gender expectations (Öttl and Behne, 2016), the aim of the present study is to investigate whether mode of acquisition affects the formation of new representations. To achieve this aim, the present study replicates an experiment in which experience-based gender expectations were induced in an explicit learning situation, but shifting the context to an implicit learning scenario.

One reason why mode of acquisition might be expected to affect how new representations are processed or stored is based on the assumption that mode of acquisition guides a learner’s attention (e.g., Marsden et al., 2013). For example, prior knowledge that a to-be-acquired artificial lexicon encodes referential gender by means of suffixation likely makes a learner consciously focus on the suffixes available in the language, but also on visual features that are likely to be informative of gender in possible referents. Thus, one might expect that in an explicit learning scenario, attention both to relevant linguistic and visual information would be enhanced from the onset of learning. If, on the other hand, a learner lacks explicit knowledge about the structure underlying the artificial language, detecting and correlating the linguistic and visual regularities necessarily requires at least one additional learning step. In the latter case, relevant gender information could potentially remain undetected in one or both modalities, resulting in less attention to gender, which again could have implications for the associations that are being established in the learning process. Crucially, whether implicit learning results in conscious or unconscious knowledge about the gender coding system, a learner may successfully acquire unanalyzed mappings between words and referents, i.e., forming an association between two holistic units (word and referent) without realizing that there is a systematic relationship between suffix and visual gender cues. Thus, in terms of mastering word-referent associations, similar learning outcomes are possible from both modes of acquisition, but superficial similarities may also overshadow potential differences

in processing, such as in the formation of experience-based expectations.

In a recent study in which participants acquired a gender-coded artificial language based on a suffixation system similar to that of Spanish, Öttl and Behne (2016) found that experience-based gender expectations (1) can be simulated by manipulating relative frequencies of exposure to male vs. female referents during training, (2) surface during online lexical processing, and (3) are not overridden by linguistic cues. In this study, participants acquired associations between spoken pseudowords on the one hand and visual referents on the other. Whereas pseudowords were marked for gender by means of suffixation, visual referents represented novel imaginary figures whose facial features were gendered by means of stereotypically masculine or feminine traits, making gender an integral feature of the referents. To induce experience-based expectations about a referent’s gender, relative frequencies of exposure to male vs. female words and their associated referents were manipulated during training. Thus, each wordstem would be unambiguously associated with a figure’s overall features (color, texture, and shape), and would also be more likely to appear with either the female or the male suffix (or both would be equally likely) and the figure of the corresponding gender. Results showed that participants were faster at identifying those referents that were consistent with the induced expectation than those that were not. This finding indicates that that learners did not solely rely on the unambiguous information that was available from the linguistic input, but also developed experience based expectations about referential gender that surfaced during subsequent processing. While this study provides insights into the formation and activation of experience-based gender expectations on the lexical level, it was based on an explicit learning situation in which learners were initially informed about the gender coding system underlying the materials to be acquired, and it is not clear to which extent the observed effects depended on this prior knowledge. The present study replicates this experiment in order to assess whether experience based gender expectations can also be induced in an implicit learning scenario, and thereby provides a more stringent simulation, as the category of interest (gender) is not highlighted to the learners. Since overt attention to gender is likely to be lower in an implicit learning situation, one possibility is that gender expectations are not induced. A less obvious possibility is that implicit learning results in stronger gender expectations. The rationale underlying the latter possibility is that in an explicit learning scenario, learners know that the information that is needed to identify the gender of a referent is encoded in the suffix, and they may therefore be less sensitive to probabilistic information that potentially contradicts the linguistic information.

Both in the original experiment and the current replication, participants acquire word-referent associations by immediate feedback. Images of four possible referents are presented on a screen, a pseudoword is presented auditorily, and participants have to select one of the available candidates. Of particular importance for the current experiment, other lines of research have demonstrated that learners may use cross-situational

statistics to detect word-referent mappings, i.e., tracking the co-occurrence of potential referents for a given word across trials can be sufficient to establish the correct mappings over time, even in the absence of direct evidence, see e.g., Yu and Smith (2007). In addition to the direct mapping that is enabled by immediate feedback, learners may also apply additional learning mechanisms. For example, as participants become increasingly familiar with the stimuli, recognizing one or more of the distractors means that these can be eliminated from the possible candidates (for a discussion of the mutual exclusivity principle, see Markman and Wachtel, 1988). Crucially, while word-referent associations may be acquired as one-to-one mappings, the structure of the artificial language implies that parallel mappings exist between word stems and figures on the one hand, and between suffixes and gender on the other, whether the participant becomes consciously aware of this or not. If, and when, these component-based mappings are detected, they potentially provide learners with top-down information that can boost acquisition. The present study does not investigate these learning mechanisms *per se*, but see e.g., Koehne and Crocker (2014) for a recent study on the interplay between such learning mechanisms.

If experience-based expectations can be detected in an implicit learning scenario, we expect participants to be quicker at recognizing referents whose gender is consistent with an induced expectation relative to referents whose gender is inconsistent with an induced expectation. Additionally, if such effects are replicated, any observed differences in strength will inform how mode of acquisition affects the resulting representation. One possibility is that stronger gender expectations are observed in an implicit learning scenario relative to an explicit learning scenario. This finding would suggest that in an explicit learning scenario, unambiguous linguistic information attenuates the impact of experience based information, even if it does not override it *per se*. On the other hand, finding weaker gender expectations in an implicit learning scenario would suggest that the explicit instructions, at least to some extent, contributed to the development of gender expectations, most likely by priming learners' awareness of gender cues in both modalities. These questions have implications for understanding the impact of different modes of acquisition, but also the cognitive representation of gender information, by assessing the impact of explicit awareness of gender coding on the resulting expectations.

## 2. METHODS

### 2.1. Design

The present experiment is a replication of an experiment reported in Öttl and Behne (2016), where participants were trained in a gender-coded artificial language in which the frequency of exposure to different words and referents were manipulated to induce experience-based expectations. In the current replication, one aspect of the original experiment was modified. Rather than explicitly informing participants that they would acquire a gender-coded language, this information was withheld in order to establish an implicit learning situation. Apart from this modification and the removal of one example

slide illustrating the gender coding, the two experiments are identical, and consist 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 new word-referent associations, and (c) a post-test in which the processing of the newly acquired representations is evaluated. The different parts of the experiment, and the frequency manipulation, are outlined in more detail in Section 2.4.

### 2.2. Participants

Twenty native speakers of Norwegian (10 male, mean age = 23.1,  $SD = 2.8$ ) 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, and gave their informed consent by signing a form that had been approved by the Data Protection Official for Research at Norwegian Universities (NSD).

### 2.3. Materials

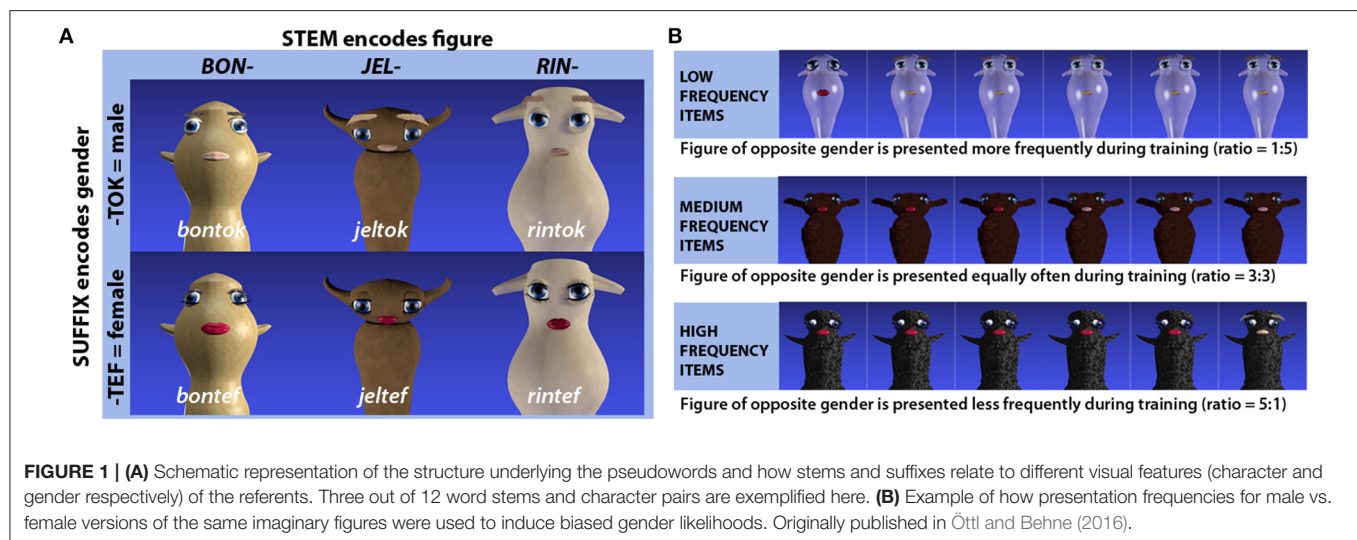
All materials used in the experiment are identical to those used in Öttl and Behne (2016), where the development of the stimuli is described in more detail.

#### 2.3.1. Auditory Stimuli

The artificial lexicon was designed to encode gender through suffixation and consisted of 24 pseudowords. These were made up of 12 pseudoword stems, which were paired with two different pseudosuffixes (“-tef” and “-tok”) (see **Figure 1**). Structurally, the pseudowords were made up of two syllables, each of which consisted of a consonant-vowel-consonant sequence. The 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 and Weenink, 2017). As fine acoustic-phonetic detail can be actively used to predict upcoming information during online processing at the lexical level (Salverda et al., 2003), participants could theoretically exploit acoustic-phonetic cues from the word stem to predict whether it would end in “-tef” or “-tok.” To ensure that the suffix could not be predicted from the stem, the 24 original recorded tokens were therefore 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 and Weenink, 2017) that were used interchangeably throughout the experiment. The average duration of pseudowords was 865 ms ( $SD = 71$ ). The timepoint at which gender information become available can be identified as the onset of the vowel in the second syllable. Both for words ending in -tef and words ending in -tok this occurred 440 ms after word onset ( $SD = 46$  and  $SD = 47$  respectively).

#### 2.3.2. Visual Stimuli

Imaginary figures were designed to provide referents for the artificial language outlined above, and could be either male or



female. This image set was entirely symmetric in the sense that it was based on 12 base figures without any cues to gender. These base figures were distinguished in terms of overall shape, color and surface texture (e.g., shiny, furry, matte), and for each base figure, a male and a female version was created. While female figures had red lips and long eye-lashes, male figures had lighter but short eye-lashes, slightly smaller pink lips and bushy eyebrows. All gender cues were thus local features of the facial region, in contrast to the global features distinguishing the different base figures from each other. Images were created using Blender 3D modeling software, version 2.60 (Blender Foundation, 2012).

### 2.3.3. Sound–Image Associations

Each of the 12 word stems was consistently linked to one of the 12 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

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), and the experiment duration was approximately 1 h, including an additional 15–30

min for calibration, questionnaires and debriefing. Participants were informed about the overall structure of the experiment (i.e., that it contained a pre-test, training blocks and a post-test) prior to participation, but were naïve to critical aspects of the experiment (i.e., to the gender coding and the frequency manipulation).

### 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. Each trial began with a gaze contingent fixation cross in the center of the display. As soon as this had been fixated for 500 ms, four images (two male and two female figures) appeared on the display. Five hundred ms later, a pseudoword corresponding to one of these was presented over the headphones. Once a response had been made, a gray frame appeared around the selected image to indicate that the response had been registered. Five hundred ms later, all images were removed from the display. If no image was selected within 4,500 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 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. These constraints on the randomizations were implemented in order to make it more difficult for participant to detect the structure of the materials. 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 part.

### 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. Five hundred ms later (or 4,500 ms after word onset, if no image had been selected), the incorrect images were removed from the display, while the correct image remained until the pseudoword had been repeated over the headphones. Randomization procedures were identical to the pre-test. 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 frequency groups. For one third of the image pairs, the ratio of presentation was 1:5 for male vs. female versions, resulting in the male version becoming a low frequency item and the female version a high frequency item. For another third of the items, the presentation ratio was reversed to a presentation ratio of 5:1 for male vs. female realizations. For the final third, male and female realizations were presented equally often (medium frequency items). Each training block was followed by feedback on the percent of the responses which were correct, and a 30-s break.

#### 2.4.3. Post-test:

The trial structure and randomization procedures were identical to the pre-test. With respect to the visual displays 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, referred to as *no competitor trials*, was identical to the pre-test, and always contained four unrelated images. In these trials, 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 referred to as *target competitor trials*, contained an image associated with the same base figure as the target word, but of the opposite gender. 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, *distractor competitor trials*, featured two distractors constituting an image pair. The latter trial type was included to prevent participants from adopting a response strategy according to which the mere presence of an image pair would indicate that the target was among the pair. In the post-test, all words/figures appeared twice as a target in each of the three trial types, regardless of presentation frequencies during training.

Based on pre-testing and the results reported in Öttl and Behne (2016), we expect participants to acquire the 24 words within the training blocks provided. If participants also successfully acquire the gender-coding underlying the stimulus materials, post-test scores for target-competitor trials (that require explicit gender identification) should be similar to no-competitor and distractor-competitor trials. Regarding response times for successfully acquired word-referent pairs, we expect

participants to be quicker at recognizing high-frequency items than low-frequency items, provided that gender information is readily available. Crucially, we also expect the gazedata to provide information on how quickly gender information becomes available.

#### 2.4.4. Statistical Procedures

All analyses 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 Supplementary Materials. In line with current recommendations (Barr et al., 2013), maximal random effects structure was used, i.e., in addition to random by-subject intercepts, random by-subject slopes were included for the fixed effects being tested. Also, the contribution of random slopes to the model fit was assessed using the same forward testing approach as described above. The inclusion of random slopes was warranted for all models. 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 trials where no image pairs were present, and this estimate can be directly compared to the adjustments required for target competitor trials and distractor competitor trials. Correspondingly, when frequency is included as a fixed effect, the intercept represents low frequency items and direct comparisons to medium and high frequency items are available from the model estimates. When both effects are included in the same model, the intercept represents the estimate for low frequency 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 proportions instead of log likelihoods and milliseconds instead of log transformed milliseconds. To test differences between the factor levels that cannot be read directly from the model (i.e., between medium and high frequency 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. RESULTS AND DISCUSSION

Section 3.1 below presents the learning curves to evaluate participants' overall performance throughout the training. Sections 3.2 through 3.4 focus on results from the post-test only: accuracy data (3.2), response times (3.3), and gazedata (3.4). Interim discussions are included in the respective sections.

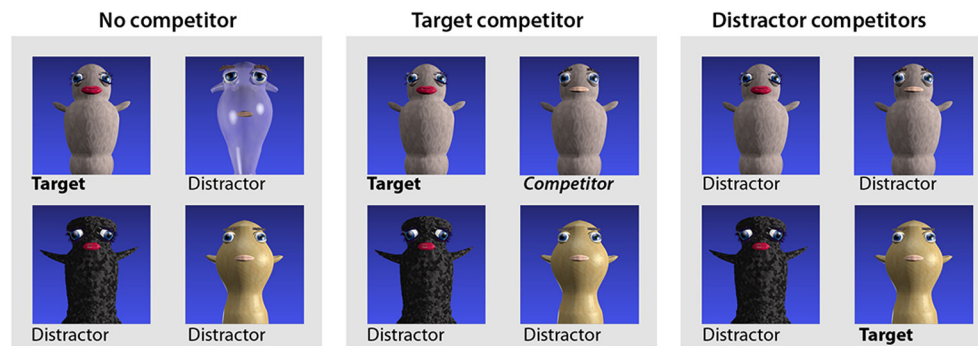


**FIGURE 2 |** Different trial types used in the post-test. Originally published in Öttl and Behne (2016).

### 3.1. Learning Curves

As shown in **Figure 3A**, participants started at chance levels in the pre-test, and successfully acquired the artificial language within the provided training blocks. In the last training block (Training5), only two participants scored below 90% correct (at 88.9 and 83.3%). Since they demonstrate a similar progression in learning as the other participants, they are included for further analysis.

**Figure 3B** presents the proportion of gender-coding errors. During training, a gender-coding error indicates that a participant in addition to not mastering stem-character links fails to acknowledge suffix-gender links. A low proportion of gender-coding errors would indicate that the gender-coding system had been deciphered. **Figure 3B** shows that participants did not immediately detect the gender-coding. That these errors decrease in parallel with overall learning increases implies that participants could in principle be forming instance-based mappings. If this were the case, we would not expect any learning transfer to occur between the two versions of a given character type, as these would be treated as independent instances. To follow up on this possibility, participants' mean accuracy at the second exposure to a given character type in the first training block was analyzed. When this second exposure constituted a true repetition (e.g., the exposure to "bontef" had been preceded by one exposure to "bontef" earlier in the training), participants scored at 44% correct ( $SD = 18.7$ ). When the exposure constituted a false repetition (i.e., the exposure to "bontef" had been preceded by exposure to "bontok," but "bontef" had not yet been encountered in training) participants scored at 48% correct ( $SD = 26.6$ ). The difference between the two scores was not significant [ $t(19) = 0.499$ ,  $p = 0.623$ ]. Since participants are also likely to use cross-situational information (as discussed in the introduction), and the randomization of trials did not control for the time-point at which these second exposures constituted true or false repetitions, a parallel analysis of trial numbers was conducted. The average trial number was 19 ( $SD = 3$ ) for true repetitions and 21 ( $SD = 7$ ) for false repetitions, but this difference was not significant [ $t(19) = 1.156$ ,  $p = 0.262$ ]. Taken together, these results suggest that participants needed some time to detect the gender-coding system, but that they nevertheless were sensitive to the referential overlap between male and female versions of a

character early on. This is consistent with research demonstrating that learners can track one-to-one and one-to-many mappings in parallel, particularly when it comes to natural categories (Gangwani et al., 2010).

### 3.2. Post-test Accuracy

As outlined in Section 2.4, the post-test includes target competitor trials, where participants are required to actively distinguish the male and the female realization of the same character. An elevated error rate in these trials would suggest that participants primarily relied on information from the word stem, and therefore likely experienced targets and competitors as ambiguous referents. An alternative scenario would be that a decrease in performance is indicative of switching costs or more general confusion or surprise at the new trial type. Importantly, the performance in distractor competitor trials is likely to be informative, as these are visually identical to the target competitor trials, however without requiring a gender distinction to be made. Potentially, the errors committed in target competitor trials may also reflect effects of frequency, in which case accuracy is expected to be highest for high frequency items and/or lowest for low frequency items.

To analyze the effects of trial type and frequency on accuracy, a binomial linear mixed effects analysis was performed. Including a fixed effect for trial type (model A) led to a significant improvement over the null model [ $\chi^2(2) = 7.8$ ,  $p < 0.05$ ]. To assess the effect of frequency, two additional models were tested: one in which a fixed effect for frequency was added to model A (model B), and one that also included its interaction term with trial type (model C). Neither of these led to significant improvements over model A [model B:  $\chi^2(2) = 2.0$ ,  $p = 0.367$ ], [model C:  $\chi^2(6) = 7.59$ ,  $p = 0.270$ ].

Fixed effects estimates from model A, which provided the best fit for the data, are presented in **Table 1**. For trials where no image pair was present, model A estimates a score of 98.4% correct (95% CI [97.1, 99.1]). In trials where a distractor competitor pair was present, this score is estimated to be 96.5%, which is significantly lower (95% CI [95.0, 97.6],  $p < 0.01$ ). Also in trials where a target competitor was present, performance is significantly worse at 94.6% correct (95% CI [88.8, 97.5],  $p < 0.001$ ). No significant difference was found between target and distractor competitor

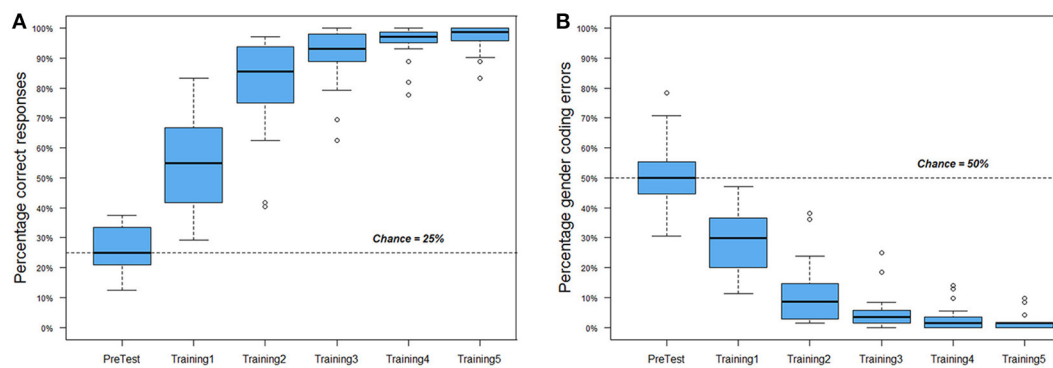


**FIGURE 3 |** Performance during training blocks. Boxes represent median values and upper and lower quartiles. **(A)** Overall accuracy. **(B)** Gender-coding errors.

**TABLE 1 |** Model estimates for accuracy data (Day 1).

| Fixed effects          | Estimate | Std. Error | z-value | Pr(<z) |
|------------------------|----------|------------|---------|--------|
| (Intercept)            | 4.089    | 0.291      | 14.05   | <0.001 |
| <b>TRIAL TYPE</b>      |          |            |         |        |
| Distractor competitors | -0.761   | 0.279      | -2.73   | <0.01  |
| Target competitor      | -1.218   | 0.354      | -3.44   | <0.001 |

**TABLE 2 |** Model estimates for response time data.

|                   | Estimate | S.E   | df   | t-value | Pr(>t) |
|-------------------|----------|-------|------|---------|--------|
| (Intercept)       | 7.543    | 0.037 | 19.1 | 203.84  | <0.001 |
| <b>TRIAL TYPE</b> |          |       |      |         |        |
| Distractor comp.  | -0.035   | 0.015 | 74.5 | -2.43   | <0.05  |
| Target comp.      | 0.076    | 0.022 | 19.1 | 3.45    | <0.01  |
| <b>FREQUENCY</b>  |          |       |      |         |        |
| Medium            | -0.050   | 0.019 | 23.0 | -2.67   | <0.05  |
| High              | -0.039   | 0.015 | 57.1 | -2.51   | <0.05  |

trials ( $p = 0.229$ ). The lower accuracy in target competitor trials suggests that the gender-coding did lead to some difficulties. However, since accuracy was also lower in distractor competitor trials, the decrease in performance cannot be fully attributed to difficulties with gender-coding, but at least partially to the new visual displays. That no effects of frequency were observed may be due to ceiling effects.

### 3.3. Response Times

In the following, only correct responses that were longer than 300 ms are analyzed (95% of the data), as earlier responses are more likely to be erroneous button presses than to reflect actual recognition (e.g., Baayen, 2008). As suggested in Baayen (2008), response times were log transformed prior to the analysis. As shown in Section 3.2, performance differed between the trial types, and this implies that removing incorrect responses affected the three trial types to different degrees: 97% of the data were analyzed for no competitor trials, 96% for distractor competitor trials, and 85% for target competitor trials.

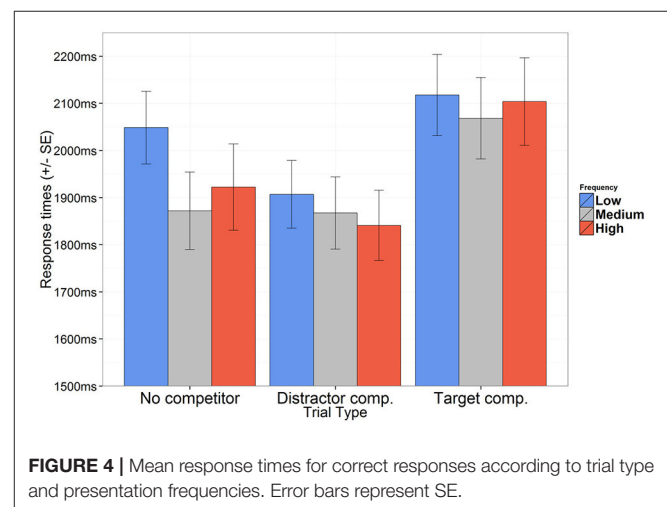


**FIGURE 4 |** Mean response times for correct responses according to trial type and presentation frequencies. Error bars represent SE.

Based on the findings in Öttl and Behne (2016), response times are expected to be longer for trials where a target competitor is present, compared to trials where none is present (partly because participants need to await auditory information from the suffix in order to identify the target). For trials where a distractor competitor pair is present, response times are expected to be shorter, since two response alternatives can be eliminated as soon as the stem has been identified. If experience based expectations can be replicated in an implicit learning situation, these are expected to surface as longer response times for low frequency items and/or shorter response times for high frequency items, reflecting relative ease of processing. The best fitting model includes fixed effects for trial type and frequency, but not their interaction term. The estimates obtained from this model are summarized in **Table 2**, and the aggregated data are presented in **Figure 4**.

The response time is estimated at 1,831 ms (95% CI[1,699, 1,974]) when no competitor is present. Relative to this, response times were significantly longer in target competitor trials (1,976 ms, 95% CI[1,835, 2,127]), and significantly shorter in distractor competitor trials (1,768 ms, 95% CI[1,645, 1,900]). Compared to the overall response time for low frequency items (1,908

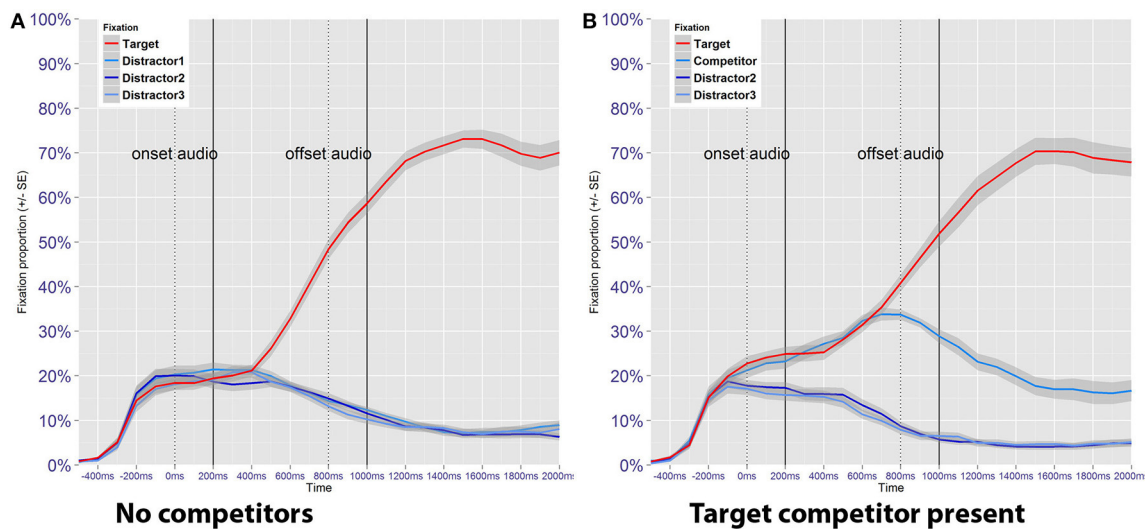


**FIGURE 5 |** Proportion of fixations toward the target image as a function of time, plotted separately for no competitor trials (A) and target competitor trials (B). Blue lines represent fixation proportions toward the three distractor images, which in (B) features a target competitor. Dotted lines represent approximate acoustic onsets and offsets for the auditory stimuli, while the solid lines include a 200 ms shift to account for the temporal lag between language processing and eye movement execution.

ms, 95% CI[1,782, 2,042]), response times were significantly shorter both for medium frequency items (1,814 ms, 95% CI[1,683, 1,955]) and high frequency items (1,836 ms, 95% CI[1,698, 1,984]). These results show that experience based expectations reported in Öttl and Behne (2016) were successfully replicated.

### 3.4. Gazedata

Gazedata 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. Each obtained gaze coordinate was classified as pertaining to one of four regions of interest (corresponding to the four image positions on the display), or as falling outside these regions. To compensate the 200 ms that are typically estimated for the planning and execution of eye movements (e.g., Matin et al., 1993), the time windows of analysis are shifted correspondingly, as is common for this paradigm (e.g., Huettig and Altmann, 2005). Two epochs of the timeline are of particular interest. On one hand, expectations based on presentation frequency may be driven by information available from the word stem, and such an effect can be expected to be detectable in the time-window defined from 200 – 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. On the other hand, expectations 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 600–1,000 ms after the onset of the word.

For an initial exploration, the proportion of fixations toward target and distractor images was calculated in time bins of 100 ms, aggregated by subject and trial type (Figure 5). Correct and

incorrect responses are included in order to reflect overall timing of stimulus events and to capture global patterns in the data. When no image pairs are present (Figure 5A), participants are equally likely to fixate either of the four images until 400 ms after the acoustic onset of the target word. From this timepoint onwards incoming auditory information is used incrementally to identify the correct image, as reflected in the increased number of fixations toward the target image. Also in line with the expectations, Figure 5B shows that fixation proportions toward the target image and the target competitor do not bifurcate until disambiguating information from the suffix becomes available. This happens approximately 600–700 ms after the acoustic onset (only after information from the first 400–500 ms of the word has been processed) which coincides with the onset of the suffix. An additional pattern apparent in Figure 5B is that fixations to the image pair is higher even before auditory information is available. This pattern was also found in distractor competitor trials (not shown in the figure), and indicates that the mere presence of an image pair attracted participants' attention.

The analysis presented in the following sections is conducted on the two time windows as previously defined, after the removal of trials with incorrect responses. Conducting separate analyses on the different time windows also allows time to be modeled as a linear predictor, which facilitates model specification, estimation and interpretation. The analysis of trials without an image pair is followed by a separate analysis of target competitor trials. These trial types were analyzed separately because the initial exploration revealed that the overall gaze patterns differ, and because target competitor trials allow for a direct comparison between low and high frequency items within the same display.

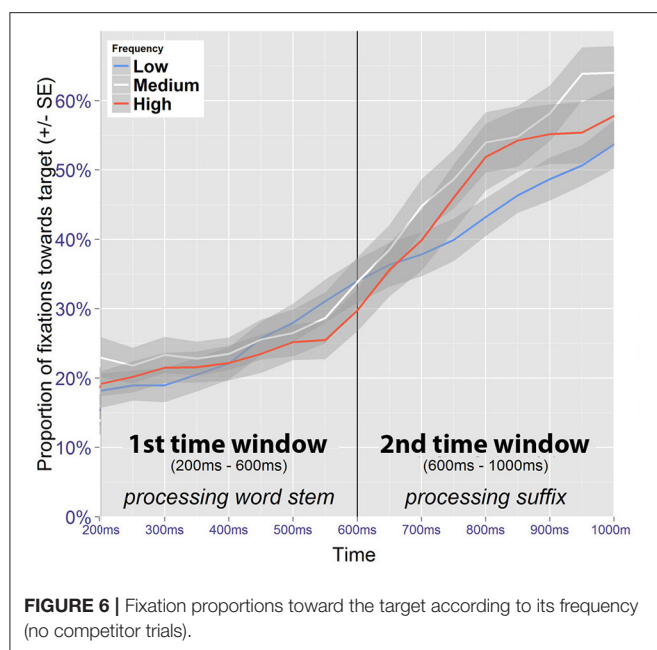


**FIGURE 6 |** Fixation proportions toward the target according to its frequency (no competitor trials).

**TABLE 3 |** Model estimates, first time window.

|                  | Estimate | S.E   | z-value | Pr(>t) |
|------------------|----------|-------|---------|--------|
| (Intercept)      | -1.848   | 0.164 | -11.28  | <0.001 |
| <b>FREQUENCY</b> |          |       |         |        |
| Medium           | 0.343    | 0.139 | 2.47    | <0.05  |
| High             | 0.182    | 0.160 | 1.14    | 0.256  |
| <b>TIME</b>      |          |       |         |        |
| Time*Medium      | -0.499   | 0.142 | -3.51   | <0.001 |
| Time*High        | -0.477   | 0.146 | -3.27   | <0.01  |

### 3.4.1. Trials without an Image Pair (No Competitor Trials)

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. In order to obtain estimates 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. **Figure 6** presents fixation proportions toward the target figure according to its presentation frequency in relation to the two time windows of interest.

#### 3.4.1.1. Word stem

For the gaze patterns in the time window corresponding to the processing of the word stem (200–600 ms after auditory onset), only target fixations are included in the analysis. To retain maximal temporal resolution, unaggregated data were used.

The best fitting model contains fixed effects of time and frequency along their interaction term (**Table 3**). At the

**TABLE 4 |** Model estimates, second time window.

|                  | Estimate | S.E   | z-value | Pr(>t) |
|------------------|----------|-------|---------|--------|
| (Intercept)      | -0.804   | 0.125 | -6.42   | <0.001 |
| <b>FREQUENCY</b> |          |       |         |        |
| Medium           | 0.068    | 0.158 | 0.43    | 0.668  |
| High             | -0.031   | 0.175 | -0.18   | 0.861  |
| <b>TIME</b>      |          |       |         |        |
| Time*Medium      | 0.472    | 0.122 | 3.87    | <0.001 |
| Time*High        | 0.485    | 0.123 | 3.93    | <0.001 |

beginning of the time-window, the fixation proportion toward low frequency targets is estimated at 13.6% (95% CI [10.3, 17.8]). This estimate is significantly higher for medium frequency targets (18.6%, 95% CI [14.5, 22.6],  $z = 2.47$ ,  $p < 0.05$ ). For high frequency targets, the estimate is also higher, but this difference is not significant (15.9%, 95% CI [12.1, 20.6],  $z = 1.14$ ,  $p = 0.256$ ). At the end of the time window, the fixation proportion toward low frequency targets is estimated at 28.6% (95% CI [22.2, 36.0]). By comparison, medium frequency targets have a fixation proportion of 25.5% (95% CI [20.8, 30.8]), which is not significantly different ( $z = -1.17$ ,  $p = 0.24$ ). For high frequency items, the estimate is 23.0% (95% CI [18.0, 28.8]), and this is marginally significant ( $z = -1.89$ ,  $p = 0.059$ ).

Although these results suggest frequency-based information to be effective during online processing of the stem, their interpretation is not straight forward. Crucially, the increased amount of fixations is observed at the beginning of the time window, and is therefore likely to be a spurious effect, as a language driven effect would be expected to increase (or at least to be sustained) within the time window.

#### 3.4.1.2. Suffix

To investigate possible frequency effects coinciding with the processing of the suffix, fixations toward the target image were analyzed in the time window ranging from 600–1,000 ms after the onset of the target word.

Significant effects were found for frequency and time, along their interaction (**Table 4**). Low frequency images received 31% of the fixations (95% CI [26, 36]) at the beginning of the time window. This is not significantly different from the estimate obtained for medium frequency images (32%, 95% CI [26, 40],  $z = 0.43$ ,  $p = 0.668$ ), nor for high frequency images (30%, 95% CI [24, 38],  $z = -0.18$ ,  $p = 0.861$ ). In contrast, at the end of the time window the fixation proportion for low frequency images is estimated at 50% (95% CI [44, 56]). This estimate is significantly higher for medium frequency images (63%, 95% CI [56, 70],  $z = 3.43$ ,  $p < 0.001$ ), and also for high frequency images (61%, 95% CI [53, 69],  $z = 2.6$ ,  $p < 0.01$ ).

The results suggest a processing disadvantage for stimuli that were inconsistent with the induced expectations. That the effect of frequency is not manifest at the beginning of the time window, but rather emerges within it, further suggests that this effect is at least partially driven by the processing of the suffix, as opposed to being a continuation of the potentially spurious effect observed

in the preceding time window. A possible explanation would be that participants first used auditory information from the stem to identify the correct figure, but were expecting a different suffix, and therefore became uncertain at this timepoint.

### 3.4.2. Trials Featuring a Target Competitor (Target Competitor Trials)

The presence of an image pair in a singular display offers the opportunity to investigate fixations toward low and high frequency images as paired observations. If an image pair for which a bias has been introduced is being fixated, the participant is either looking at the figure consistent with the induced expectation, or she is looking at the figure that is inconsistent with this expectation. Crucially, participants are also either looking at the male or the female member of the pair, and may show systematic choices in which of the two is fixated first. Acknowledging these dependencies, only fixations toward the image pair were analyzed. The adequacy of this approach is supported by the fact that the presence of an image pair generally attracts participants' attention (**Figure 5**), which results in the number of observations being higher than for no competitor trials.

Data were recoded to define fixations toward the male image as the dependent variable, and the analyses investigated the effects of time on the one hand and of the induced gender bias on the other. For trials where the male figure had a high presentation frequency during training, the gender bias is referred to as being male, while for trials where the male figure had had a low presentation frequency during training, the gender bias is referred to as being female. **Figure 7** shows the fixation proportions toward the male member of the pair for the two time windows of interest, according to the induced gender bias. The global pattern suggests that participants tended to fixate the male figure first.

#### 3.4.2.1. Word stem

The first time window (200–600 ms after onset of the word) was analyzed for trials featuring a target competitor. Only correct trials where a bias had been introduced were included in the analysis. The best fitting model contains fixed effects for time and bias, along their interaction (**Table 5**). When there was a male bias, male images received 65.3% of the fixations at the beginning of the time window. When there was a female bias, fixations toward male images was lower at 59.9%, but the difference was not significant ( $z = 1.11$ ,  $p = 0.267$ ). At the end of the time window, the fixation proportion toward the male image had dropped to 47.5% when there was a male bias, and to 31.9% when there was a female bias. This difference was significant ( $z = 3.22$ ,  $p < 0.01$ ). Thus, despite the initial and bias-independent preference for fixating the male image at the beginning of the time window, when there was a female bias, there was also a clear preference for fixating the female image at the end of the time window. This suggests that probabilistic information available from the word stem was indeed used to predict which of the two images was going to be the referent.

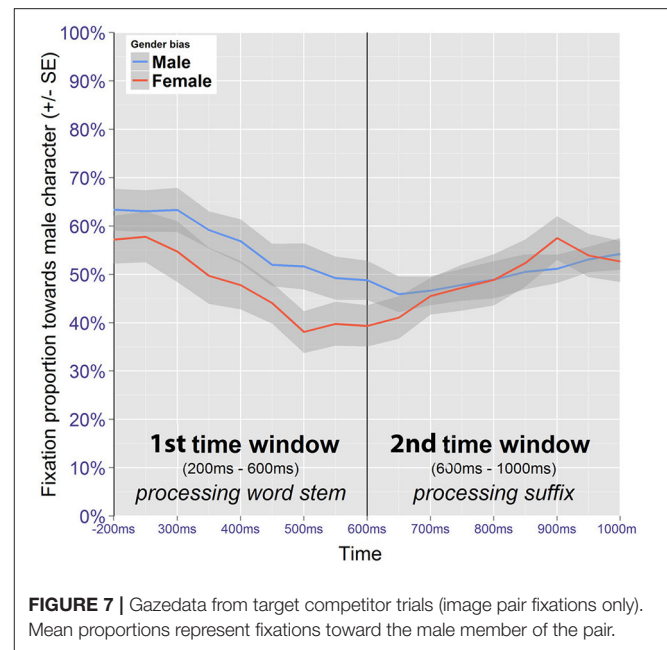


**FIGURE 7 |** Gazedata from target competitor trials (image pair fixations only). Mean proportions represent fixations toward the male member of the pair.

**TABLE 5 |** Model estimates, first time window.

|                    | Estimate | S.E   | z-value | Pr(> z ) |
|--------------------|----------|-------|---------|----------|
| (Intercept)        | 0.633    | 0.205 | 3.09    | <0.01    |
| <b>GENDER BIAS</b> |          |       |         |          |
| Female             | -0.232   | 0.209 | -1.11   | 0.267    |
| <b>TIME</b>        | -0.732   | 0.248 | -2.95   | <0.01    |
| Time*Female        | -0.428   | 0.182 | -2.36   | <0.05    |

**TABLE 6 |** Model estimates, second time window.

|                    | Estimate | S.E   | z-value | Pr(> z ) |
|--------------------|----------|-------|---------|----------|
| (Intercept)        | -0.239   | 0.167 | -1.43   | 0.153    |
| <b>GENDER BIAS</b> |          |       |         |          |
| Female             | -0.260   | 0.200 | -1.30   | 0.194    |
| <b>TIME</b>        | 0.403    | 0.195 | 2.07    | <0.05    |
| Time*Female        | 0.476    | 0.155 | 3.07    | <0.01    |

#### 3.4.2.2. Suffix

The same analysis strategy was used on the time window that coincides with the processing of the suffix (600–1,000 ms after the onset).

Male images receive 44.1% of the fixations at the beginning of the time window when the bias is toward the male image (**Table 6**). When the bias is toward the female image, the estimate is lower at 37.8%, but not significantly different ( $z = -1.30$ ,  $p = 0.194$ ). At the end of the time window, the fixation proportion toward the male image has increased to 54.1% when there is a male bias and to 59.4% when there is a female bias, but again the difference is not significant ( $z = 0.197$ ,  $p = 0.273$ ).

The gaze patterns show that at the beginning of the second time window, male figures receive more attention when there is a male bias, compared to when there is a female bias, and that this difference is attenuated over time. Nevertheless, this difference does not reach significance. Regardless of its origin, the attenuation of the effect within the time window suggests that disambiguating gender information from the suffix was indeed used to identify the correct image.

In summary, the patterns observed in the preceding analyses indicate that frequency information affects online processing. When no competitor was present, a preference for looking at the high frequency image was detected during the processing of the suffix. When a target competitor was present, the preference for the high frequency image was detected during processing of the stem.

## 4. GENERAL DISCUSSION

The present study addresses the emergence of experience-based gender expectations in an implicit learning situation, using an artificial language paradigm. The results show that during the acquisition of a miniature artificial language consisting of gender-marked pseudowords and associated visual referents, participants track the frequency of exposure to male vs. female realizations of words and referents, and thereby build up gender expectations inherent to the new representations. These expectations surface during subsequent processing, even if the gender-coding system underlying the materials was acquired implicitly. On a global level, the results of the current experiment replicate findings reported in an explicit learning scenario (Öttl and Behne, 2016). Whereas accuracy and response time data to some extent indicate that similar representations were acquired in the current and the replicated experiment, suggesting that experience-based expectations are robust against manner of acquisition, the investigation of gender coding errors and eyetracking data reveal a slightly more complex picture.

In terms of accuracy during acquisition, participants started off at 25% correct in the pre-test. Lacking explicit knowledge about the gender-coding system underlying the material, the four available candidate referents were equally plausible, and therefore 25% correct mirrors chance level performance. By the third training block, performance approaches ceiling, suggesting that the items have been successfully acquired. However, bearing in mind that gender information was redundant during the training blocks, as targets could be identified by the wordstems alone, the high accuracy at this stage likely overestimates participants' knowledge of the gender-coding system. Evidence that knowledge about the gender coding is indeed lower than the performance in the training blocks suggests can be found in the results from the post-test, where performance differs according to trial type. When an image pair is present (i.e., in target competitor and distractor competitor trials), performance is significantly worse relative to when it is not (no competitor trials). Accuracy being lower when an image pair is present suggests that the gender coding was at least to some extent problematic for the participants. At the same

time, the fact that performance was lower both in target competitor trials and in distractor competitor trials indicates that it is not necessarily due to gender coding alone, since in the distractor competitor trials the drop in performance can only be attributed to the visual presence of an image pair among the distractors. The mere presence of an image pair in the visual display is therefore also likely to be a contributing factor to the lower performance in target competitor trials. The difference between target competitor and distractor competitor trials (94.6% vs. 96.5% correct respectively) did not reach significance. However, bearing in mind that performance was at 98.4% correct in trials without a competitor, the contrast in performance strongly suggests that the gender coding required to resolve target competitor trials makes this trial type even more difficult.

Turning to the post-test response time data, which offer a window on the newly formed representations that goes beyond mere accuracy, shorter response times were found for both medium and high probability items relative to low probability items, indicating that experience-based gender expectations were successfully induced. Recognizing a referent whose gender is consistent with an induced expectation (or, in the case of medium probability items, items for which no gender expectation has been induced) is quicker than recognizing one whose gender is inconsistent with the induced expectation. No indication was found that the facilitation could be gradual depending on the probability, as a facilitation of 94 ms was observed for medium frequency items and one of 72 ms for high frequency items, a difference that did not reach significance. Additional evidence that experience-based gender expectations affected online aspects of processing was found in the eyetracking data. Overall, the gaze patterns were consistent with the expectations for the visual world paradigm to the extent that with incoming auditory information, participants became increasingly more likely to fixate the target referent, indicating that the newly acquired lexicon was processed similarly to natural words (see e.g., Dahan et al., 2001). Gender expectations were found to surface during online processing of the words, with increased fixations toward target referents whose gender was consistent with an induced expectation relative to targets whose gender was inconsistent with an induced expectation. When a target competitor was present, this effect was observed during the processing of the stem, and it was attenuated during the processing of the suffix. Also in no competitor trials, this effect was observed, but only in the time window corresponding to the processing of the suffix.

Globally, the patterns outlined above are very similar to the patterns reported for an explicit learning scenario in Öttl and Behne (2016), with only a few exceptions. The most evident difference between results from the two experiments is to be found in the accuracy data. In both experiments, participants started at chance level performance. In the current experiment, chance level lies at 25% correct, as the participants had to select one out of four possible referents. In Öttl and Behne (2016) however, participants were explicitly instructed about the gender-coding system underlying the artificial language and were also aware of the visual gender cues, and therefore chance level performance in that experiment is at 50%, as the

display always featured two male and two female characters. Though in both experiments performance approaches ceiling in the third training block, suggesting that the word-referent associations are mastered to a similar degree regardless of the mode of acquisition, an examination of the performance in the post-test reveals that this initial assessment is too superficial. While in an explicit learning situation, participants performed at ceiling regardless of trial type, the presence of an image pair led to a weakened performance in an implicit learning situation. Thus, even if performance is very high in both experiments, the gender coding system cannot be said to be perfectly mastered in the implicit learning situation employed in the current experiment. When it comes to the induced gender expectations as reflected in the response time data, the patterns found in the current experiment are very similar to those reported in Öttl and Behne (2016). Both in an implicit and in an explicit learning situation, a significant facilitation is found for both medium and high probability items relative to low probability items. Whereas Öttl and Behne (2016) find the facilitation to be by 27 and 96 ms for medium and high probability items respectively, such a gradual effect is not found in the present experiment. Nevertheless, the overall extent of the facilitation is highly consistent across the two experiments, approaching a maximum contrast of 100 ms. We do not have a plausible explanation for why the effect is not gradual in an implicit learning scenario, but acknowledge that in neither experiment was a significant difference found between medium and high probability items. Even if the gender expectations are not perfectly mirrored in the current replication experiment, it is noteworthy that they appear to a similar extent, particularly if seen in light of the general differences found for gender coding *per se*.

Additional evidence that an implicit learning scenario led to difficulties with gender coding relative to an explicit learning scenario can be found by comparing the response times according to trial type across the two experiments. In the current experiment, the presence of a distractor competitor led to a facilitation of 63 ms when compared to trials in which no competitors were present, which is similar to the facilitation of 56 ms reported for an explicit learning scenario (Öttl and Behne, 2016). This facilitation is likely due to the fact that the image pair can be quickly eliminated from consideration, reducing the number of available candidates. For target competitor trials on the other hand, the findings are more divergent, with the current experiment yielding a delay of 145 ms, compared to only 75 ms in Öttl and Behne (2016). As argued in Section 3.3, finding longer response times for target competitor trials must be partially attributed to the fact that auditory information from the suffix is required to resolve reference, and it is not clear whether the 75 ms delay reported in Öttl and Behne (2016) may be fully attributed to this aspect, or if it is also partly due to target competitor trial being experienced as more difficult as well. In either case, a longer delay for target competitor trials in the current experiment is consistent with the difficulties with gender coding observed in accuracy measures.

One limitation to the above comparison between the current and the replicated experiment is that the contrast between

implicit and explicit learning is understood as a contrast between providing and not providing participants with knowledge about the gender coding system underlying the material to be acquired. In Öttl and Behne (2016), participants were shown an example of an image pair and informed which suffix encoded which gender. Distinguishing to which extent the differences in results can be attributed to visual vs. linguistic aspects of processing is therefore not possible. The overall gaze patterns are similar across the two studies, except for the gender expectations seeming to arise somewhat earlier in the current experiment. One possible explanation for this contrast is that it is driven primarily by differences in visual attention during the inspection of the figures. Support for this view can be found in the overall gaze patterns, where Öttl and Behne (2016) report that during the processing of the suffix in no competitor trials, participants were more likely to fixate the distractor of the same gender as the target than those of the opposite gender, an indication that the suffix was used actively to guide visual attention. A corresponding effect was not detected in the current experiment.

That experience based gender expectations seem at least to some extent to be independent of overt attention to gender information provides support for the finding that with real words, gender expectations are hard or difficult to suppress (Oakhill et al., 2005). In the current experiment, participants were not instructed or in any way encouraged to pay attention to gender information, and results show that the gender-coding system was at least to some extent difficult to acquire. Nevertheless, the sensitivity to probabilistic gender information seems to come very close to that observed in an explicit learning situation. Though for real words, the timescale of acquisition and the complexity of referents and learning contexts are undeniably of a different scale than the simulations presented here, a central implication of these findings is that experience-based gender expectations reflect statistical regularities in the input, regardless of whether the categories these regularities belong to are highlighted or not.

## 5. CONCLUSION

By replicating an experiment investigating the formation of frequency based expectations in an artificial language, while changing the learning situation from an explicit to an implicit one, the present experiment contributes to understanding the impact different modes of acquisition may have on the formation of new representations. Finding that the acquisition of a gender coding system proceeds less successfully in an implicit learning scenario than in an explicit learning scenario may not be surprising in its own right, since in the latter case what it to be learned is already given away. More importantly, the finding that frequency based expectations seem to surface to similar extents in both learning scenarios indicates that certain aspects of newly formed representations are robust against manner of acquisition. This finding has implications not only for understanding differences between different modes of acquisition, but also for understanding the

cognitive representation of gender information, particularly why gender expectations are activated, even in cases where they are not relevant for discourse coherence.

## ETHICS STATEMENT

This study did not involve the collection or storage of any personal data, and did therefore not require clearance by an ethical committee. All subjects gave their written informed consent, and were free to withdraw at any stage during the study. The protocol was approved by NSD (Norwegian Centre for Research Data).

## 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.

## FUNDING

The present research was conducted within the Marie Curie Initial Training Network: Language, Cognition, & 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. The authors would also like to thank Pirita Pyykkönen-Klauck for helpful comments on the analyses.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01485/full#supplementary-material>

## REFERENCES

- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. New York, NY: Cambridge University Press.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models Using Eigen and S4*. R Package Version, 1(7). Available online at: <http://cran.r-project.org/package=lme4>
- Blender Foundation (2012). *Blender - a 3D Modelling and Rendering Package*. Amsterdam: Blender Online Community, Blender Institute.
- Boersma, P., and Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. *Glot Int.* 6.0.28, 341–345. Available online at: <http://www.praat.org/>
- 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
- Gangwani, T., Kachergis, G., and Yu, C. (2010). "Simultaneous cross-situational learning of category and object names," in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (Austin, TX), 1595–1600.
- Gygax, P. M., Garnham, A., and Doehren, S. (2016). What do true gender ratios and stereotype norms really tell us? *Front. Psychol.* 7:1036. doi: 10.3389/fpsyg.2016.01036
- 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
- Koehne, J., and Crocker, M. W. (2014). The interplay of cross-situational word learning and sentence-level constraints. *Cogn. Sci.* 39, 849–889. doi: 10.1111/cogs.12178
- 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. Available online at: <http://cran.r-project.org/package=lmerTest>
- 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
- Markman, E. M., and Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cogn. Psychol.* 20, 121–157. doi: 10.1016/0010-0285(88)90017-5
- Marsden, E., Williams, J., and Liu, X. (2013). Learning novel morphology. *Stud. Second Lang. Acquisit.* 35:619. doi: 10.1017/S0272263113000296
- Martin, 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., Gygax, 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
- Öttl, A., and Behne, D. M. (2016). Experience-based probabilities modulate expectations in a gender-coded artificial language. *Front. Psychol.* 7:1250. doi: 10.3389/fpsyg.2016.01250
- 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.
- 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
- Singer, J. D., and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. London: Oxford University Press.
- Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x

**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 © 2017 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.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership