

# BILINGUAL LANGUAGE DEVELOPMENT: THE ROLE OF DOMINANCE

EDITED BY: Cornelia Hamann, Esther Rinke and Dobrinka Genevska-Hanke  
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 978-2-88945-988-9  
DOI 10.3389/978-2-88945-988-9

## 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)



# BILINGUAL LANGUAGE DEVELOPMENT: THE ROLE OF DOMINANCE

Topic Editors:

**Cornelia Hamann**, University of Oldenburg, Germany

**Esther Rinke**, Goethe University of Frankfurt, Germany

**Dobrinka Genevska-Hanke**, University of Oldenburg, Germany



Image: Larissa Stamm

It has long been established that bilingual speakers are rarely balanced in their languages so that one language is dominant. The contributions to the Research Topic “Bilingual Language Development: The Role of Dominance” focus on the potential effects of language dominance on the competence and processing of bilinguals, covering a large variety of language combinations and domains. Important aspects of such work are the interplay of L1-maintenance/attrition and possible L2-dominance, the direction of cross-linguistic influence (CLI) or code-mixing, as well as the effects of bilingualism on cognitive development, each addressed in several contributions. However, such research presupposes a definition of dominance, which is far from being settled. This gives rise to considerable differences in the operationalization

of the concept across studies. The studies in this Research Topic present a multifaceted picture of the role of language dominance for L1-maintenance/attrition, L2-development and CLI. Though a unified story cannot emerge for such a complex subject, interesting new venues are explored including the impact of dominance shift during L1-reexposure, comparisons of different types of bilingual groups, or operationalization of dominance through experiential measures. The variety of approaches and results is in part owed to the many language combinations studied and the fact that bilingual children, adults and atypical speakers are investigated. This diversity constitutes the interest of this Research Topic.

**Citation:** Hamann, C., Rinke, E., Genevska-Hanke, D., eds. (2019). Bilingual Language Development: The Role of Dominance. Lausanne: Frontiers Media.  
doi: 10.3389/978-2-88945-988-9



# Table of Contents

- 06 Editorial: Bilingual Language Development: The Role of Dominance**  
Cornelia Hamann, Esther Rinke and Dobrinka Genevska-Hanke
- 09 What Machine Learning can Tell us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks**  
Lina Abed Ibrahim and István Fekete
- 41 Vocabulary, Metalinguistic Awareness and Language Dominance Among Bilingual Preschool Children**  
Carmit Altman, Tamara Goldstein and Sharon Armon-Lotem
- 57 Different Outcomes in the Acquisition of Residual V2 and Do-Support in Three Norwegian-English Bilinguals: Cross-Linguistic Influence, Dominance and Structural Ambiguity**  
Merete Anderssen and Kristine Bentzen
- 75 Multilingual Competence Influences Answering Strategies in Italian–German Speakers**  
Irene Caloi, Adriana Belletti and Cecilia Poletto
- 89 Age of Onset and Dominance in the Choice of Subject Anaphoric Devices: Comparing Natives and Near-Natives of Two Null-Subject Languages**  
Elisa Di Domenico and Ioli Baroncini
- 104 Input Dominance and Development of Home Language in Russian-German Bilinguals**  
Natalia Gagarina and Annegret Klassert
- 118 Bilingualism in a Case of the Non-fluent/agrammatic Variant of Primary Progressive Aphasia**  
Nomiki Karpathiou, John Papatriantafyllou and Maria Kambanaros
- 134 First Language Attrition and Dominance: Same Same or Different?**  
Barbara Köpke and Dobrinka Genevska-Hanke
- 150 Morpho-Syntactic Abilities of Unbalanced Bilingual Children: A Closer Look at the Weaker Language**  
Natalia Meir
- 167 Bilinguals' Sensitivity to Grammatical Gender Cues in Russian: The Role of Cumulative Input, Proficiency, and Dominance**  
Natalia Mitrofanova, Yulia Rodina, Olga Urek and Marit Westergaard
- 187 Case Marking in Hindi as the Weaker Language**  
Silvina Montrul, Archana Bhatia, Rakesh Bhatt and Vandana Puri
- 206 Language Dominance and Cognitive Flexibility in French–English Bilingual Children**  
Elena Nicoladis, Dorothea Hui and Sandra A. Wiebe
- 217 Language Dominance Affects Bilingual Performance and Processing Outcomes in Adulthood**  
Eloi Puig-Mayenco, Ian Cunnings, Fatih Bayram, David Miller, Susagna Tubau and Jason Rothman

- 233** *Predictors of Language Dominance: An Integrated Analysis of First Language Attrition and Second Language Acquisition in Late Bilinguals*  
Monika S. Schmid and Gülsen Yilmaz
- 252** *The Age Factor Revisited: Timing in Acquisition Interacts With Age of Onset in Bilingual Acquisition*  
Petra Schulz and Angela Grimm
- 270** *Experiential Measures can be Used as a Proxy for Language Dominance in Bilingual Language Acquisition Research*  
Sharon Unsworth, Vicky Chondrogianni and Barbora Skarabela



# Editorial: Bilingual Language Development: The Role of Dominance

Cornelia Hamann<sup>1\*</sup>, Esther Rinke<sup>2\*</sup> and Dobrinka Genevska-Hanke<sup>1\*</sup>

<sup>1</sup> Institute for English and American Studies, University of Oldenburg, Oldenburg, Germany, <sup>2</sup> Institute of Romance Languages and Literatures, Goethe-Universität Frankfurt am Main, Frankfurt, Germany

**Keywords:** bilingualism, language acquisition, dominance, L1 attrition, dominance operationalization

## Editorial on the Research Topic

### Bilingual Language Development: The Role of Dominance

It has long been established that bilingual speakers are rarely balanced in their languages so that one language is dominant. The contributions to Bilingual Language Development: The Role of Dominance focus on the potential effects of language dominance on the competence and processing of bilinguals, covering a large variety of language combinations and domains. Important aspects of such work are the interplay of L1-maintenance/attrition and possible L2-dominance, the direction of cross-linguistic influence (CLI) or code-mixing, as well as the effects of bilingualism on cognitive development, each addressed in several contributions. However, such research presupposes a definition of dominance, which is far from being settled. This gives rise to considerable differences in the operationalization of the concept across studies. Among other factors, many researchers use proficiency to determine dominance (Genesee et al., 1995), others exposure and use (Argyri and Sorace, 2007), or environmental language (Polinsky, 2008, but see Schmeißer et al., 2015). Recently, a trend developed toward an integrated perspective (Birdsong, 2018). In their overview, Cantone et al. (2008) argue for a definition combining experiential with performance factors (see also Montrul, 2015 and Silva-Corvalán and Treffers-Daller, 2016) while Bedore et al. (2012) demonstrate that experiential measures and relative proficiency are related. More specifically, Unsworth (2016) argues that experiential variables predict performance/proficiency, so that relative amount of exposure and use can serve as a proxy for language dominance. Complementing this work, Unsworth et al. estimate relative exposure and use with a parental questionnaire and obtain performance measures through spontaneous speech production as well as standardized vocabulary measures. Their results indicate that language use is a stronger predictor of proficiency than exposure, pointing to its importance for dominance.

These findings are crucial for research on language impairments. For developmental language disorder (DLD) it is recommended to test children in both languages or at least in their dominant language and adjust norms according to dominance (Thordardottir, 2015) while therapy for aphasia should be conducted in the dominant language. Three studies in this collection address language impairments. Abed Ibrahim and Fekete investigate bilingual children with German as early-L2. They operationalize dominance through experiential factors and investigate performance in two repetition tasks. Using partitioning around medoids, which does not rely on a priori group assignment, they show that dominance influences performance in typical bilinguals without negatively affecting diagnostic accuracy. Their findings suggest that testing in the majority language does not necessarily disadvantage bilingual children. Meir compares the morpho-syntactic abilities of four groups of Russian-German bilingual children: children with a weaker language, balanced bilinguals and children dominant in that language, as well as bilingual children with DLD. While error patterns are the same across typical bilingual groups, unbalanced bilinguals used complex

## OPEN ACCESS

### Edited and reviewed by:

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

### \*Correspondence:

Cornelia Hamann  
cornelia.hamann@uni-oldenburg.de  
Esther Rinke  
esther.rinke@em.uni-frankfurt.de  
Dobrinka Genevska-Hanke  
dobrinka.genevska.hanke@  
uni-oldenburg.de

### Specialty section:

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

**Received:** 18 April 2019

**Accepted:** 24 April 2019

**Published:** 21 May 2019

### Citation:

Hamann C, Rinke E and  
Genevska-Hanke D (2019) Editorial:  
Bilingual Language Development: The  
Role of Dominance.  
Front. Psychol. 10:1064.  
doi: 10.3389/fpsyg.2019.01064



syntax, relying on resources from the dominant language. However, bilinguals with DLD simplified structures. This suggests that the non-dominant language of unbalanced typical bilinguals may be delayed, not deviant, whereas acquisition patterns of bilinguals with DLD are distinct, resembling those of monolingual children with DLD. While there is an extensive body of research on bilingual children with DLD, studies on bilingual patients with neurodegenerative diseases are much rarer. Karpathiou et al. thus provide a valuable study of the factors determining language preservation. They show that lexical and grammatical abilities of a bilingual with L1-Greek and late L2-French are impaired in both languages, with better preservation of the dominant language.

CLI and language mixing provide more evidence for the importance of a dominance definition. Investigating the acquisition of residual V2-structures in English by three Norwegian-English simultaneous bilingual children, Andersson and Bentzen find different patterns of CLI, but argue that these differences cannot be attributed to dominance. They also discuss whether dominance determines the direction of mixing (Genesee et al., 1995) arguing that dominance, when determined by proficiency and experiential measures, does not affect mixing. Also focussing on CLI, Puig-Mayenco et al. examine the role of dominance for the competence of children with L1-Spanish/early-L2 Catalan, and children with L1-Catalan/early-L2 Spanish. Focusing on one property of Catalan, the occurrence of pre-verbal Negative Concord Items (NCIs) not licit in Spanish, and one of Spanish more restricted in Catalan, Differential Object Marking (DOM), the authors observe CLI for DOM but not for NCIs. L1-Spanish/L2-Catalan speakers over-accept DOM in Catalan, suggesting that their dominance in Spanish influences the directionality of CLI, which the performance of the L1-Catalan/L2-Spanish group confirms.

It has been observed that language dominance may shift during development, and Gagarina and Klassert investigate the L1-competence of bilingual children in such a dominance-shift context: the systematic exposure to L2 in kindergarten/preschool. They examine experiential factors, such as input provided by the nuclear family, as predictors for lexical and (verbal/nominal) morphological development in L1 after change of input dominance. Interestingly, age, gender, and L2-AoO all differently impact these domains. Importantly, verbal inflection proves to be more robust than case inflection. The latter finding ties in with the longitudinal study of Schulz and Grimm investigating experiential measures and, crucially, timing in acquisition as factors influencing development of the majority language. The study compares simultaneous to early-L2 bilingual children and shows that while age of onset affects early-acquired phenomena such as subject-verb-agreement, this is not the case for late-acquired phenomena such as case marking. Clearly, timing modulates age effects. Dominance, determined by experiential measures, does not play a role for early or late acquired phenomena in simultaneous bilinguals.

Language dominance and attrition are interrelated, established with similar measures and influenced by the same factors so that they might represent two stages of the same phenomenon as suggested by Köpke and Genevskaya-Hanke.

Among several studies on L1-attrition, Schmid and Yilmaz offer an integrated perspective investigating the role of various experiential and proficiency predictors of language dominance in four migrant populations. Focusing on dominance shifts and factors facilitating L1-attrition/maintenance, they suggest that different aspects of bilingualism affect language development in different ways. For L1, measures of informal language use play a role in determining whether a bilingual is a good maintainer, while success in L2-acquisition depends heavily on personal factors such as the educational level. Similarly, Montrul et al. find effects of exposure and use on the knowledge of Hindi-case marking in Hindi-English bilinguals with different dominance patterns: balanced bilinguals in India outperform unbalanced L2 and heritage speakers in the US, who show instability in their production/knowledge of the Hindi case system. In contrast, the study of Mitrofanova et al. finds that a combination of experiential and proficiency measures is the best predictor for different attrition/maintenance patterns. Studying the acquisition of gender in Russian by Russian-Norwegian bilingual children in Norway they show that bilingual and monolingual children were sensitive to phonological gender cues, albeit to different degrees.

Whereas the previous studies reveal specific domains as problematic for L2-acquisition or L1-maintenance in different experiential situations, Caloi et al. provide a broader view. They compare adult heritage speakers of Italian, late L2-learners with L1-Italian, and Italian monolinguals as to the strategy employed when answering new information questions. Monolinguals prefer a Verb-Subject-structure, whereas heritage and L2-speakers behave alike, but different from monolinguals, in opting for Subject-Verb-answers. These group comparisons lead the authors to remind the reader that a “bilingual speaker is not two monolinguals in one” (Grosjean, 1989)—the grammatical features of L1 are well-mastered and it is the richer experience which leads to a different, a wider, not an attrited or incomplete system.

Apart from the language domains studied so far, the use of null- and overt subjects has long been in the focus of research on the effects of language dominance on L2-development and L1-maintenance, and two of the contributions address this domain. Di Domenico and Baroncini look at the role of dominance and age of onset on the choice of overt and null pronominal subjects in native and near-native speakers of Greek and Italian. The results reveal age of onset as a factor: near-natives overuse overt pronouns, simultaneous bilinguals do not. Interestingly, effects of dominance are found for null pronouns and lexical DPs but not for overt pronouns.

Turning to psycholinguistic aspects of bilingualism and language dominance, Köpke and Genevskaya-Hanke define dominance as the relative accessibility of each of the languages of a bilingual for language processing. Taking late L1-attrition and dominance to represent different stages of a continuum, they investigate knowledge of pronominal subjects (see also Di Domenico and Baroncini) in a speaker of (pro-drop) L1-Bulgarian and (non-pro-drop) L2-German, who had late L2-onset and fairly long residency in Germany. The authors argue that attrition of a highly-entrenched L1 affects language processing not underlying representations, and that it does so

temporarily only, disappearing fast after a limited re-exposure to L1-input. This opens the question if re-exposure to Italian would have changed the performance of the bilinguals in the study by Caloi et al..

Regarding notions of language accessibility, an advantage in cognitive flexibility has been attributed to bilinguals (Bialystok, 2005), but studies have not always reached the same conclusions. Nicoladis et al. contribute to the discussion whether the need to access each language depending on context accounts for flexibility by investigating whether balanced bilinguals have an advantage over other bilinguals. Study of French-English bilingual children shows that none of the experiential or proficiency measures predicts cognitive flexibility. Also focusing on cognitive aspects, Altman et al. investigate the influence of dominance on metalinguistic awareness, MA, and whether MA mediated by dominance influences vocabulary size. Crucially, only the Hebrew vocabulary size of the Russian-dominant children is affected by MA. They rely on their fast mapping

abilities to expand vocabulary, which, the authors argue, reflects their level of vocabulary development in the societal language.

The studies in this volume present a multifaceted picture of the role of language dominance for L1-maintenance/attrition, L2-development and CLI. Though a unified story cannot emerge for such a complex subject, interesting new venues are explored including the impact of dominance shift during L1-re-exposure, comparisons of different types of bilingual groups, or operationalization of dominance through experiential measures. The variety of approaches and results is in part owed to the many language combinations studied and the fact that bilingual children, adults and atypical speakers are investigated. This diversity constitutes the interest of this volume.

## AUTHOR CONTRIBUTIONS

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

## REFERENCES

- Argyri, E., and Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Biling. Lang. Cogn.* 10, 79–99. doi: 10.1017/S1366728906002835
- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Green, K., et al. (2012). The measure matters: language dominance profiles across measures in Spanish-English bilingual children. *Biling. Lang. Cogn.* 15, 616–629. doi: 10.1017/S1366728912000090
- Bialystok, E. (2005). “Consequences of bilingualism for cognitive development,” in *Handbook of Bilingualism. Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. de Groot (Oxford: Oxford University Press), 417–433.
- Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Front. Psychol.* 9:81. doi: 10.3389/fpsyg.2018.00081
- Cantone, K. F., Kupisch, T., Müller, N., and Schmitz, K. (2008). Rethinking language dominance in bilingual children. *Linguistische Berichte* 215, 307–343. Available online at: <https://buske.de/rethinking-language-dominance-in-bilingual-children.html>.
- Genesee, F., Nicoladis, E., and Paradis, J. (1995). Language differentiation in early bilingual development. *J. Child Lang.* 22, 611–631.
- Grosjean, F. (1989). Neurolinguists, beware! the bilingual is not two monolinguals in one person. *Brain Lang.* 36, 3–15.
- Montrul, S. (2015). “Dominance and proficiency in early and late bilingualism,” in *Language Dominance in Bilinguals*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35.
- Polinsky, M. (2008). Gender under incomplete acquisition: heritage speakers’ knowledge of noun categorization. *Herit. Lang.* J.6, 40–71.
- Schmeißer, A., Hager, M., Gil, L. A., Jansen, V., Geveler, J., Eichler, N., et al. (2015). “Related but different: the two concepts of language dominance and language proficiency,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 36–65.
- Silva-Corvalán, C., and Treffers-Daller, J. (2016). *Language Dominance in Bilinguals. Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press.
- 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. doi: 10.21832/9781783093137-014
- Unsworth, S. (2016). “Amount of exposure as a proxy for dominance in bilingual language acquisition,” in *Language Dominance in Bilinguals*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173.

**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 © 2019 Hamann, Rinke and Genevskaya-Hanke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# What Machine Learning Can Tell Us About the Role of Language Dominance in the Diagnostic Accuracy of German LITMUS Non-word and Sentence Repetition Tasks

Lina Abed Ibrahim<sup>1\*</sup> and István Fekete<sup>2</sup>

<sup>1</sup> Department of English, University of Oldenburg, Oldenburg, Germany, <sup>2</sup> Department of Dutch, University of Oldenburg, Oldenburg, Germany

## OPEN ACCESS

### Edited by:

Esther Rinke,  
Goethe-Universität Frankfurt am Main,  
Germany

### Reviewed by:

Angela Grimm,  
Goethe-Universität Frankfurt am Main,  
Germany

Sharon Armon-Lotem,  
Bar-Ilan University, Israel

Elma Blom,  
Utrecht University, Netherlands

### \*Correspondence:

Lina Abed Ibrahim  
lina.abed.ibrahim@uni-oldenburg.de

### Specialty section:

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

**Received:** 05 June 2018

**Accepted:** 20 December 2018

**Published:** 30 January 2019

### Citation:

Abed Ibrahim L and Fekete I  
(2019) What Machine Learning Can  
Tell Us About the Role of Language  
Dominance in the Diagnostic  
Accuracy of German LITMUS  
Non-word and Sentence Repetition  
Tasks. *Front. Psychol.* 9:2757.  
doi: 10.3389/fpsyg.2018.02757

The present study investigates the performance of 21 monolingual and 56 bilingual children aged 5;6–9;0 on German LITMUS-sentence-repetition (SRT; Hamann et al., 2013) and non-word-repetition-tasks (NVRT; Grimm et al., 2014), which were constructed in accordance with the **LITMUS**-principles (**L**anguage **I**mpairment **T**esting in **M**ultilingual **S**ettings; Armon-Lotem et al., 2015). Both tasks incorporate phonologically and syntactically complex structures shown to be cross-linguistically challenging for children with Specific Language Impairment (SLI) and aim at minimizing bias against bilingual children while still being indicative of the presence of language impairment across language combinations (see Marinis and Armon-Lotem, 2015; for sentence-repetition; Chiat, 2015 for non-word-repetition). Given the great variability in bilingual language exposure and the potential effect of language experience on language performance in bilingual children, we examined whether background variables related to bilingualism, particularly, the degree language dominance as measured by relative amount of use and exposure, could compromise the diagnostic accuracy of the German LITMUS-SRT and NVRT. We further investigated whether a combination of the two tasks provides better diagnostic accuracy and helps avoid cases of misdiagnosis. To address this, we used an unsupervised machine learning algorithm, the Partitioning-Around-Medoids (PAM, Kaufman and Rousseeuw, 2009), for deriving a clinical category for the children as  $\pm$  language-impaired based on their performance scores on SRT and NVRT (in isolation and combined) while withholding information about their clinical status based on standardized assessment in their first (home language, L1) and second language (societal language, L2). Subsequently, we calculated diagnostic accuracy and used regression analysis to investigate which background variables (age of onset, length of exposure, degree of language dominance, socio-economic-status, and risk factors for SLI) best explained clinical-group-membership yielded from the PAM-analysis based on the children's NVRT and SRT performance scores. Results show that although



language-dominance clearly influences the performance of bilingual typically developing children, especially in the SRT, the diagnostic accuracy of the tools is not compromised by language dominance: while risk factors for SLI were significant predictors for clinical group membership in all models, language dominance did not contribute at all to explaining clinical cluster membership as typically developing or SLI based on any of the combinations of the SRT and NWRT variables. Additionally, results confirm that a combination of SRT scored by correct target structure and the structurally more complex language-dependent part of the NWRT yields better diagnostic accuracy than single measures and is only sensitive to risk factors for SLI and not to dominance levels or SES.

**Keywords:** bilingualism, specific language impairment, sentence repetition, non-word repetition, language dominance, k-medoid clustering algorithm, unsupervised learning, conditional inference trees

## INTRODUCTION

Recent research in language disorders has focused on problems of language assessment and the identification of what is currently referred to in the literature as Developmental Language Disorder (DLD, see Bishop et al., 2017) or Specific Language Impairment (SLI<sup>1</sup>) in bilingual children. The latter term refers to a disorder in the development of language in the absence of auditory, cognitive, sensory-motor, neurological, or socio-emotional deficits (Leonard, 1998, 2014). A challenge constantly facing clinicians is to determine whether a bilingual child's poor performance on language tasks in the societal language (second language-L2) is due to an inborn language impairment (LI) or to insufficient exposure to the L2 (cf. Armon-Lotem et al., 2015; Marinis et al., 2017).

A major contributor to the diagnostic difficulties of SLI is the heterogeneity of children with SLI, who constitute a group with diverse linguistic profiles and deficits of varying severity across language components (Crutchley et al., 1997; Conti-Ramsden et al., 2001; Friedmann and Novogrodsky, 2011; Leonard, 2014 among others). For many children with SLI, deficits in the area of morphosyntax (grammatical morphology and syntactic structure) stand out (Leonard, 2007; Marinis and van der Lely, 2007; Marinis, 2011). On the one hand, certain complex syntactic structures with linguistic operations involving dependencies such as syntactic movement (e.g., Wh-questions) and embedding (e.g., relative clauses), have been shown to be cross-linguistically problematic for children with SLI (Jakubowicz et al., 1998; van der Lely, 1998; Friedmann and Novogrodsky, 2011; Jakubowicz, 2011; Hamann and Tuller, 2014; Hamann et al., 2017). On the other hand, SLI may manifest itself differently depending on the language being acquired so that clinical markers vary across languages.

Problems of language-impaired children are not restricted to the morphosyntactic domain, albeit being most deficient. Various studies have shown that children with SLI also evince

deficits in the area of phonology. These children lag behind their age matched peers in the acquisition of consonants and are particularly sensitive to phonological complexity such as consonant clusters (Gallon et al., 2007; Ferré et al., 2015; dos Santos and Ferré, 2018; Grimm and Hübner, in press), coda position (Tamburelli and Jones, 2013) and syllabic position in the foot (Bortolini and Leonard, 2000). As a coping strategy, consonant clusters are often reduced or even avoided (Bortolini and Leonard, 2000; Orsolini et al., 2001; Marshall et al., 2003). Although morphosyntactic and phonological deficits are more commonly reported in the literature (Leonard, 2014), children with SLI also have deficient lexical retrieval abilities, which are not only delayed but also qualitatively different from those of children with typical language development (Novogrodsky and Kreiser, 2015). A number of studies have further shown that children with SLI exhibit deficits in the interface between syntax-semantics and pragmatics, e.g., universal quantification, telicity and exhaustivity in Wh-questions (Roeper, 2004; Schulz and Roeper, 2011). Even though children with SLI often present different combinations of the deficits, Friedmann and Novogrodsky (2008, p. 214) point to the existence of “selective impairments in one module of language, and not in others.” Accordingly, it is possible “to identify subgroups within SLI with selective deficits in various language modules: syntax [grammatical/syntactic-SLI], lexicon [lexical-SLI], phonology [phonological-SLI] and pragmatics [pragmatic-SLI]” (ibid., p. 214).

Aside from the aforementioned language deficits, a large body of research has identified deficits in phonological short-term memory, as indicated by poor performance on repeating non-words with a length of two to four syllables as a special weakness in children with SLI (Gathercole and Baddeley, 1990; Archibald and Gathercole, 2006; Gathercole, 2006; for a meta-analysis see Graf Estes et al., 2007). Although deficits in phonological short-term memory and certain aspects of grammar involving grammatical computational aspects<sup>2</sup> such as verbal morphology and syntactic comprehension often co-occur in children with SLI, evidence from a twin study by Bishop et al.

<sup>1</sup>We are aware of the recent consensus on using the term “Developmental Language Disorder-DLD” for unexplained language impairment in the absence of primary deficits. Nevertheless, we chose to refer to this disorder as SLI in this paper for the sake of continuity with much of the existing literature on bilingual SLI and our own collaborative research within the Franco-German project “BiLaD.”

<sup>2</sup>Linguistic operations such as recursion and hierarchical non-local dependencies between grammatical elements (van der Lely, 2005, p. 53).

(2006) has shown that despite being significantly heritable, the two vulnerable areas were separable. While some children displayed deficits in both areas, other children displayed deficits in one but not the other, suggesting that they are not “different manifestations of the same underlying deficit” (Leonard, 2014, p. 19).

Apart from diagnostic difficulties caused by the heterogeneity of the disorder, identifying LI in bilingual children is made far more complex by the great variability in their (typical) language development, which is influenced by a multitude of child internal and external factors (Paradis, 2011; Hamann, 2012). The latter include age of onset (AoO) of systematic (sustained) exposure to the second language (L2), length of exposure (LoE), quantity and quality of linguistic input (poor or enriched), L1-L2 typological proximity, status of the home language (high prestige, minority, or heritage language), and socioeconomic status (SES). The interplay of these factors makes it notoriously difficult to establish what is typical for bilingual language development (Tuller et al., 2018). Depending on the timing of exposure, bilingual children could be classified as simultaneous ( $AoO < 3$ ), early ( $3 \leq AoO < 4$ ) or late ( $AoO \geq 4$ ) sequential child bilinguals (also referred to as child L2, Meisel, 2009). Even in simultaneous bilingual language acquisition, bilingual children “have their input space divided” (Paradis and Genesee, 1996, p. 9) and are likely to receive less exposure to each language, on average, than monolingual age peers acquiring the respective languages. As a result, bilingual children often develop unbalanced command of their two languages, i.e., their linguistic abilities are unevenly distributed both within and across language domains at a given age (e.g., Döpke, 2000; Yip and Matthews, 2006; Kohnert, 2010). The language with the more advanced state of development within the process of language acquisition (Deuchar and Muntz, 2003; Genesee and Nicoladis, 2007; Gathercole, 2016) or the language to which the child receives more exposure on a regular basis (Pearson et al., 1997) is commonly described as the dominant (stronger) language as opposed to the weaker or non-dominant one (see also Meisel, 2007). In this sense, dominance is associated with language exposure/use (Grosjean, 2016) and/or with the degree of proficiency in either language (Petersen, 1988; Deuchar and Muntz, 2003; Genesee and Nicoladis, 2007). In the present study, we adopt Argyri and Sorace’s (2007, p. 83) definition of dominance as “the language in which the bilingual child obtains more input on a regular basis” (see also Grosjean, 2010). Language dominance can also shift over time due to changes in patterns of use and exposure resulting from “changes in family structure, child-care arrangements, schooling, or place of residence” (Paradis, 2010: p. 652). For example, in case of early sequential child bilinguals, who start acquiring the societal (second language L2) while their home language (first language, L1) is still at an early developmental stage, a change in the degree of dominance is frequently observed with schooling (cf. Flores, 2015; de Houwer and Bornstein, 2016). Diagnostic problems particularly occur when bilingual children are solely assessed using monolingual norm-referenced tests in the majority/societal language, which might still be their weaker, i.e., non-dominant language at the time

of assessment. In many cases, performance below monolingual average, especially on standardized measures for vocabulary and morphosyntax, is taken as evidence for LI leading to overdiagnosis with SLI (Bedore and Peña, 2008; Grimm and Schulz, 2014).

In addition to the aforementioned quantitative performance differences, a growing body of research has shown that the developmental trajectory of bilingual child language acquisition may show (persistent) delays (Tuller et al., 2015; Paradis et al., 2016) or temporary overlap with that of monolingual children with SLI (MoSLI), particularly in the area of morphosyntax (see Paradis, 2010 for an overview). The overlap in linguistic error patterns of bilingual typically developing children (BiTD) and error patterns serving as diagnostic markers for SLI in a particular language, e.g., extended use of infinitives in English (Rice and Wexler, 1996), object clitic omission in French (Paradis et al., 2003; Paradis, 2010; Hamann, 2012) and problems with SVA combined with the use of infinitives and verb placement errors in German (Clahsen, 1991; Hamann et al., 1998; Rothweiler et al., 2012) complicates the diagnosis of SLI in bilingual children. The delayed or deviant linguistic development of a bilingual child may be erroneously ascribed to bilingualism (underdiagnosis), while a child L2 learner may be overdiagnosed with SLI if such deficits are viewed as a token for SLI (Genesee et al., 2004; Grimm and Schulz, 2014; Armon-Lotem and de Jong, 2015), which could have costly consequences for the child and the society (Zurer-Pearson, 2010).

To avoid cases of misdiagnosis, it has been recommended to evaluate a bilingual child at least in her dominant language (Fredman, 2006) and ideally in both of her languages (American Speech-Language-Hearing Association [ASHA], 2004; Royal College of Speech and Language Therapists Specific Interest Group in Bilingualism [RCSLT], 2007; International Association of Logopedics and Phoniatrics [IALP], 2011), as genuine LI affects both. However, L1-assessment is often not feasible due to the lack of standardized language tests for (bilingual) children in their L1. Even if available, results may be unreliable due to incomplete L1-acquisition and/or L1-attrition, which are often reported for heritage language speakers (Montrul, 2008; Benmamoun et al., 2013). Not to mention that evaluation in two languages is time-consuming and that some of the immigrant L1 varieties undergo language change as a result of contact with the majority/societal language (L2), e.g., Immigrant Turkish in Germany (see Schroeder and Dollnick, 2013; Chilla and Şan, 2017). Hamann and Abed Ibrahim (2017) showed that even when dominance-adjusted bilingual cut-off criteria (Thordardottir, 2015) were applied to the standardized L1 tests, more than a quarter of the L1-dominant children in their sample were classified as SLI by the L1-tests. The fact that the latter children performed within aged-expectations on the L2-tests albeit being dominant in their heritage language questions the applicability of L1 tests in heritage contexts (even with norm adjustments) and suggests that direct assessment measures in the L2 are more reliable for identifying LI in bilingual populations, especially in case of heritage language speakers. This in turn makes it crucial to develop reliable tools that

could disentangle effects of bilingualism and LI in bilingual contexts.

## The LITMUS Tools for Bilingual Language Assessment

In an attempt to cope with the diagnostic challenges in bilingual populations, a battery of tools was designed during COST Action IS0804 “Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment” according to a set of linguistic principles that allow cross-linguistic comparability. These tools aim at minimizing the effect of factors related to bilingualism, so that SLI can be reliably identified in bilingual children with different language combinations. The latter tools are known as the LITMUS tools (Language Impairment Testing in Multilingual Settings, see Armon-Lotem et al., 2015), among which are sentence repetition (SRTs) and non-word repetition tasks (NWRs) and the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015). The latter was developed for gathering background information on factors related to bilingualism as well as information about risk factors for SLI. Such information is invaluable for the interpretation of performance results on linguistic tasks. In the current study, we concentrate on sentence repetition and non-word repetition (NWR) since they have been shown to reliably identify SLI in monolinguals (Conti-Ramsden et al., 2001) and to be less reliant on prior language experience than other language measures in bilinguals, e.g., receptive vocabulary (Chiat et al., 2013; Thordardottir and Brandeker, 2013). Depending on their construction, SRTs and NWRs can be designed to not only assess (phonological) working memory (Archibald and Gathercole, 2006), but also the command of syntactic and phonological representations/derivations (see Polišenská et al., 2015 for sentence-repetition; Gallon et al., 2007 for non-word-repetition). Such linguistic representations/derivations, especially their complexity, have been shown to crucially influence performance in these tasks (e.g., Ferré et al., 2012; Friedmann et al., 2015) so that it has been argued that they are not mere measures of working memory (Vinther, 2002; Polišenská et al., 2015). Because of this versatility, they are ideal for targeting language-specific (LS) as well as cross-linguistically challenging syntactic/phonological structures while minimizing avoidance strategies (see Hamann et al., 2017 for SRT).

Sentence repetition taps morphosyntactic abilities as recalling a sentence involves processing of the incoming input string, analysis and reconstruction thereof, especially when the sentences are long enough to prevent mere phonological reiteration (Baddeley, 2000; Marinis and Armon-Lotem, 2015). Furthermore, compared to other types of tasks, it is less constrained by pragmatic and discourse factors (Polišenská et al., 2015; Hamann et al., 2017), and is thus often used in clinical assessment as a measure of sentence-level abilities. The German LITMUS-SRT (Hamann et al., 2013) under investigation here was constructed according to the LITMUS principles (Marinis and Armon-Lotem, 2015) and builds on the notion of linguistic computational complexity. Within the generative framework, computational complexity can be determined by the number

and nature (e.g., merge vs. movement, distance of dependencies, and depth of embedding) of syntactic operations necessary for deriving a syntactic structure (Gibson, 1998; Jakubowicz, 2005; Hamann et al., 2007; Jakubowicz and Tuller, 2008; Friedmann et al., 2009). Children with atypical language acquisition are proposed to have a greater deficit on constructions with a higher degree of computational complexity, as the latter are more taxing to working memory capacities (Chomsky, 2005; Hamann et al., 2007; Jakubowicz and Tuller, 2008). A particular difficulty for children with SLI has been reported for structures involving movement along with intervening elements between the source of the moved constituent and its landing site, e.g., object *Which*-questions and object relative clauses with a lexical subject (Rizzi, 2004; Friedmann et al., 2015). Unlike the problems encountered by children with SLI, bilingual children with typical language development (BiTD) might struggle with vocabulary and uninterpretable features, i.e., grammatical features lacking semantic content like number agreement on the verb (Tsimpli and Dimitrakopoulou, 2007), or might even avoid complexity (Tuller et al., 2015). They are; however, assumed to have an intact language faculty and WM. Thus, having been acquired in the L1, syntactic operations such as recursion, embedding and movement do not have to be acquired again and should not be problematic for them given sufficient exposure to the L2 (Roeper, 2011). Accordingly, the German LITMUS-SRT incorporates a set of syntactically complex, i.e., computationally more demanding structures identified as difficult for children with SLI cross-linguistically in addition to a set of structures reported to be challenging for German MoSLI children such as topicalization and the sentence bracket, which represent crucial milestones in the acquisition of German word-order properties. The complex structures involve computational operations like syntactic movement (measured, for example by number of overt movement operations), in particular *Wh*-movement, i.e., fronting of interrogative or relative pronouns (Hamann et al., 1998; van der Lely, 1998; Marinis and van der Lely, 2007; Jakubowicz, 2011), and/or clausal embedding, e.g., relative clauses (Friedmann and Novogrodsky, 2011; Hamann and Tuller, 2014; Scheidenes and Tuller, 2018).

It has been recently shown that SRTs eliciting structures involving the latter operations can be reliably used to tease apart typically developing bilingual children from monolingual and bilingual children with SLI, not only in bilingual but also in bilingual settings (e.g., Armon-Lotem and Meir, 2016; Meir et al., 2016, 2017 for LITMUS-SRT in Russian and Hebrew; de Almeida et al., 2017; Fleckstein et al., 2018 for French; Lein et al., 2016; Abed Ibrahim and Hamann, 2017; Hamann et al., 2017; Hamann and Abed Ibrahim, 2017 for German; Theodorou et al. (2017) for Cypriot-Greek; see also Marinis et al., 2017 for an overview). In particular, Armon-Lotem and Meir (2016) showed that although the highest level of diagnostic accuracy can be achieved using a combination of SRTs in the child's L1/Russian and L2/Hebrew (applying bilingual cut-offs), good diagnostic accuracy can still be achieved if SRT is only administered in the societal language (L2-Hebrew). In the same vein, Abed Ibrahim et al. (2018) and Chilla et al. (in press) looked into the potential influence of L1-L2 typological differences on the performance of bilingual children



with Arabic, Portuguese, and Turkish as L1 on German LITMUS-SRT. L1-influence surfaced neither in the overall performance nor in the performance on the individual structures included in the task or in the expected L1-driven error patterns confirming the applicability of the task to bilingual children with diverse L1-backgrounds. It should be; however, noted that most of the studies on LITMUS-SRT report lower-cut-off scores separating TD from SLI in the bilingual groups, and that the task can only be used to assess bilinguals who had at least 12 months of exposure to the L2 (see Tuller et al., 2018).

Non-word repetition belongs to the core assessment measures used for diagnosing LI and has been identified as a reliable clinical marker of SLI in monolingual children (Conti-Ramsden et al., 2001; Gathercole, 2006). An advantage of NWR over other language measures is that it is less affected by prior knowledge of vocabulary and morphosyntax (Thordardottir and Brandeker, 2013; Chiat, 2015) and counts as a relatively culturally fair measure, which could be used for the assessment of children with diverse linguistic and socio-economic backgrounds (Engel et al., 2008; Chiat and Polišenská, 2016). As such, NWR tasks offer promising tools for the identification of SLI especially in bilingual children with limited exposure to the L2.

Measured by increasing numbers of syllables, NWR has traditionally been used to assess phonological working memory (Archibald and Gathercole, 2007; Coady and Evans, 2008). However, the ability to repeat non-words does not only rely on phonological working memory but also requires phonological skills like speech perception, phonological encoding, storage and retrieval of phonological representations, phonological assembly and articulation, which also relate to the capacity of learning new words (Gathercole, 2006). Each of these skills can be deficient in language-impaired children (Coady and Evans, 2008; Marshall, 2014). Recent studies have shown that children with SLI are not only sensitive to the amount of phonological material, i.e., number of syllables in the non-words, but also to phonological complexity such as the presence of consonant clusters, which comprise a particular source of difficulty for children with (phonological) SLI in many languages (Barlow, 2001; Gallon et al., 2007; Marshall and van der Lely, 2009; Ferré et al., 2012; Tamburelli and Jones, 2013; Leonhard, 2014).

Designing an NWRT that identifies LI in bilingual children without disadvantaging those with less experience with the L2 is not straightforward. Despite being less reliant on LS knowledge, there is substantial evidence that performance on NWR (both within and across languages) is affected by the characteristics of the non-words such as word-likeness, length, complexity, prosodic structure, phonotactic probability, and neighborhood density. For instance, children are found to perform significantly better on non-words that are more wordlike, carry LS stress patterns, contain LS-morphemes or have higher phonotactic probability (Jones et al., 2010; Messer et al., 2010; Leclercq et al., 2013; for an overview see Chiat, 2015). These findings imply that “experience and knowledge of lexical phonology contribute to NWR” (Chiat and Polišenská, 2016), which, depending on the nature of the non-words, is generally shown to relate to vocabulary size in monolingual (Gathercole, 2006) and bilingual children (e.g., Engel de Abreu et al., 2013).

Departing from that, different LITMUS-NWRTs manipulating factors shown to influence performance on NWRTs such as length, prosody and/or syllable complexity were constructed within the COST IS0804 framework for NWR (see Chiat, 2015 for details).

Similar to the LITMUS *Crosslinguistic* (Quasi-Universal) NWR test (CL-NWRT, Chiat, 2015), the German LITMUS-NWRT (Grimm et al., 2014) was constructed parallel to the French LITMUS-NWRT (dos Santos and Ferré, 2018) within the COST Action IS0804 framework for NWR tests. Unlike the CL-NWRT, e.g., the Dutch Quasi-Universal NWRT (Boerma et al., 2015; Boerma and Blom, 2017), which primarily tests phonological short-term memory and comprises phonologically simple non-words compatible with the phonological properties of any language, the German LITMUS-NWRT was devised to tap more directly into phonological abilities by focusing on phonological complexity. The latter was found to be a promising marker for assessing phonological impairment (Marshall et al., 2002; Ferré et al., 2012; for German, see Ott et al., 2006). LITMUS-NWRTs of this type systematically vary segmental (articulatory difficulty), syllabic (presence or absence of clusters) and sequential complexity (types of consonant and syllable sequences) combining them into non-words of increasing phonological complexity. At the same time, LS phonological properties are controlled as far as possible to avoid penalizing bilingual children. In order to limit effects of lexical knowledge, the non-words were constructed to be maximally distinct from real words in the target language (German) and were created using elementary blocks (segments and syllables) that are cross-linguistically well-attested (Maddieson et al., 2011). In line with the COST Action IS0804 framework (Chiat, 2015), the latter blocks were combined and manipulated in two sets, a set of phonologically complex items with phonological properties common in most of the world's languages (the quasi language-independent part, LI\_part), and an additional set of items containing the same building blocks of the LI\_part in addition to the extrametrical /s/ as a complexity variable specific to German and some other languages (the language dependent<sup>3</sup> part, LD\_part). The maximum non-word length is limited to three syllables in both parts in order to minimize working memory load, which could undermine the effect of phonological complexity. Various studies reported negative effects of language specific properties of the NWRTs on performance of bilingual children resulting in insufficient diagnostic accuracy, e.g., Kohnert et al. (2006), Windsor et al. (2010), Boerma et al. (2015), and Armon-Lotem and Meir (2016). However, since the construction of the LD\_Part in the German LITMUS-NWRT varies considerably from other LS NWRTs (see section “The German LITMUS Non-word Repetition Task”), bilingual children are not expected to be disadvantaged by the LD\_part of this particular task. Although they might encounter more difficulties with the LD items, both monolingual and bilingual children with SLI are

<sup>3</sup>Here, language-dependency is viewed as “an abstract phonological property rather than a lexical or sub-lexical property” (Grimm and Hübner, in press).

anticipated to disproportionately struggle with the structurally more complex LD items since both SLI groups are assumed to have similar underlying deficits (Paradis et al., 2011a,b). Indeed, studies by Ferré et al. (2015), dos Santos and Ferré (2018), Grimm and Hübner (in press), as well as Abed Ibrahim and Hamann (2017) have pointed to the fact that the structurally more complex LD\_part of the NWRT did not disadvantage the BiTD children, who performed on par with their monolingual peers. On the contrary, compared to the LI part, the gap between SLI and TD was larger for the LD\_part leading to better diagnostic accuracy in both monolingual and bilingual populations. These results corroborate that phonological complexity is vulnerable to phonological deficits not only in monolingual but also in bilingual children.

Several recent studies (e.g., Armon-Lotem and Meir, 2016; Meir et al., 2016; Meir and Armon-Lotem, 2017; Boerma and Blom, 2017; Tuller et al., 2018; Chilla et al., in press) investigated the diagnostic potential and impact of different variables related to bilingualism on the performance in LITMUS-SRTs and NWRTs. Here, we report on three studies of direct relevance to the present research that were conducted within the joint German-French project (BiLaD) using similar methodology with bilingual groups (Arabic/Portuguese/Turkish as L1) in Germany and France, who vary in their sociolinguistic settings. De Almeida et al. (2017) investigated the diagnostic accuracy of French LITMUS-SRT and NWRT and examined whether factors of L2 language use and exposure had an influence on the bilingual children's performance. Although both tasks significantly discriminated between SLI and TD in both monolingual and bilingual children, reduced specificity of SRT was observed for children not dominant in French. Significant correlations were found between SRT-performance and language use and dominance in the BiTD but not in the BiSLI group suggesting that dominance might be responsible for the variation observed in the BiTD group. To avoid cases of overdiagnosis and enhance diagnostic accuracy, the authors recommend combining SRT with NWRT, which did not correlate with any of the L2-exposure variables.

Tuller et al. (2018) report on direct comparisons of German and French LITMUS-NWRTs and SRTs. Their results showed good to excellent diagnostic accuracy in monolinguals, whereas the diagnostic accuracy for bilinguals was fair to good, i.e., the tasks generally distinguished bilingual children likely to be language-impaired from those likely to be typically developing. The authors further explored whether performance on the two tasks was mainly ascribed to developmental risk factors for SLI or to factors related to bilingualism. Results show that a sizable proportion of the variance in the performance of the bilinguals (BiSLI and BiTD collapsed together) in the German and French LITMUS-SRTs and NWRTs was explained by risk factors of SLI as measured by the index of Positive\_Early\_Development (see section "The LITMUS-Questionnaire for Parents of Bilingual Children" for details). Exposure and use variables such as current L2-richness accounted for additional 4% of the variance in the French-SRT and

11% of the variance in the German SRT. For the German NWRT, early L2-exposure weighed negatively to account for a further 7% of the variance. Since current L2-richness and early exposure to L2 both contribute to establishing language dominance based on the PaBiQ (see section "The LITMUS-Questionnaire for Parents of Bilingual Children"), this raises the question of whether language dominance has a negative impact on the diagnostic accuracy of the LITMUS-tools, especially on the LITMUS-SRT.

This question was further pursued in Hamann and Abed Ibrahim (2017), who used k-means cluster analysis to group bilingual children based on their performance scores on German LITMUS-SRT and NWRT as language impaired or not without access to their clinical group membership based on standardized assessment. In order to measure diagnostic accuracy, the children's k-means cluster membership based on SRT and NWRT scores was compared to the likelihood of a child to have SLI or TD based on standardized assessment in each of the child's languages (see section "Participants" for details). Whereas the sensitivity rates for both SRT (scored by identical repetition, SRT\_Id) and NWRT were excellent, the specificity rates were only suggestive, as several bilinguals were assigned to the clinical cluster based on their global NWRT and SRT\_Id scores. In line with previous studies on German LITMUS-SRT, this study showed that using the rating measure "target structure" (SRT\_Tar), which focuses on the mastery of the constructions targeted by the task, resulted in better specificity and better overall diagnostic accuracy than SRT\_Id in the bilingual groups. The individual scores of the children likely to be BiTD were plotted against language dominance for each of the tasks. While NWRT appeared to be rather unaffected by language dominance; 25% of the L1-dominant children performed below cut-off even on SRT\_Tar. Finally, the study showed that a combination of SRT and NWRT helps to avoid cases of over-identification.

Given that assessment of bilingual children is usually exclusively carried out in the societal language, the finding that dominance appears to influence the SRT performance of BiTD children, especially those dominant in their L1, raises concerns whether this task is suited for the identification of SLI in L1-dominant children when administered in their weaker language German. However, the three studies above have their limitations: in all of them, diagnostic accuracy of the tools was measured against established clinical status based on standardized evaluation in the L1 and L2, which does not take into account cases of selective impairment or problems with L1 standardized tests in heritage contexts. This, in turn, might be responsible for the reduced accuracy rates (see de Almeida et al., 2017 and Hamann and Abed Ibrahim, 2017 for a discussion). Hamann and Abed Ibrahim (2017) showed that using an alternative procedure that takes into account selective impairments and problems with L1-assessment in minority contexts minimized the slight overlap between BiTD and BiSLI and enhanced diagnostic accuracy. A further limitation is that in both of de Almeida et al. (2017) and Hamann and Abed Ibrahim (2017), dominance was not factored in as a variable into a regression analysis model and might have been confounded by other variables. Hence, the assumed influence of

dominance remains a conjuncture that needs to be statistically validated.

## The Present Study

In line with much recent research and building upon our own research, this study investigates the identification of LI in bilingual populations using sentence and nonword repetition tasks. Since both LITMUS-SRT and NWRT were designed to minimize bias against bilingual populations while being indicative of the presence or absence of LI, the following research questions emerge in the light of previous findings:

- i. Upon sufficient exposure to the L2, how robust are German LITMUS-SRT and NWRT against language dominance? Are they only sensitive to risk factors for SLI or could background variables related to bilingualism, in particular the degree of language dominance (estimated by relative amount of use of and exposure to L1/L2), compromise their diagnostic accuracy?
- ii. Since a combination of tools evaluating different aspects of language ability such as morphosyntax and phonology is recommended to acknowledge the heterogeneity within the SLI population and avoid cases of over- and underdiagnosis, does a combination of LITMUS-SRT (especially when scored by correct target structure) and NWRT yield higher accuracy rates than those estimated for each of the tasks in isolation?
- iii. Does a combination of SRT\_Tar, which evaluates the mastery of complex constructions and the phonologically more complex LD part of the NWRT provide better diagnostic accuracy for identifying SLI in (monolingual) and bilingual children than other combinations of measures?

To address these questions, we will use an unsupervised machine learning algorithm, the Partitioning Around Medoids (PAM, Kaufman and Rousseeuw, 2009) for deriving a clinical category (clustering) for the children as  $\pm$  language-impaired based on their performance scores on SRT and NWRT (in isolation and combined) while withholding information about their clinical status based on standardized assessment in L1 and L2. Subsequently, we will calculate diagnostic accuracy of the tasks (separately and combined) by verifying the goodness of the fit against the clinical groups we can establish for bilinguals by their scores in norm-referenced L1 and L2 tests (see section “Participants”), and use regression analysis to investigate which background variables (age, AoO, LoE, degree of language dominance, SES, and risk factors for SLI) best explained clinical-group-membership based on the children’s NWRT and SRT performance scores. Our premise is that if the PAM-cluster membership can be predicted by the presence of risk factors for SLI but not by any of the other background variables known to influence performance of bilingual children on language tests, particularly the degree of language dominance, then clustering of cases cuts across the SLI/TD dimension confirming that the LITMUS-SRT and NWRT are sensitive to LI and are not biased against bilingual children regardless of their language dominance.

## MATERIALS AND METHODS

### Establishing Language Dominance in Child Bilinguals

A number of methods have been put forward for measuring and operationalizing language dominance in bilingual children. These measures fall into two categories: performance-based measures and experiential-based measures (Unsworth, 2016; Unsworth et al., 2018). Estimates of language dominance obtained by performance-based measures are based on quantitative differences in proficiency measurements between the two languages of a bilingual. These measures are usually extracted from (a) spontaneous speech data, such as mean length of utterance (MLU), upper bound (UB, length of the longest utterance in a speech sample), multi-morphemic utterances (MMU), lexical diversity measures (number of different word types, verbs, and nouns) and directionality of code-mixing (see Cantone et al., 2008; Kupisch, 2008; Bedore et al., 2012 for an overview), and (b) proficiency measures based on standardized tests for vocabulary and grammar. Experiential measures, on the other hand, rely on biographical information and estimates of language use and exposure to predict dominance in bilingual children. The rationale behind the latter approach is that the (relative) proficiency of bilingual children in each of their languages is “in some sense a function of the amount of language to which they are exposed in these two languages” (Unsworth, 2016, p. 156). Accordingly, experiential variables like the relative amount of language use and exposure can be used as a predictor for the degree of bilingual language dominance.

Bedore et al. (2012), Unsworth (2016) as well as Unsworth et al. (2018) found that relative amount of exposure and use reliably predicted dominance group membership as determined by proficiency measures, confirming that relative amount of use and exposure can be used as a proxy for language dominance in bilingual children. For the purposes of the present study and building upon the findings of Bedore et al. (2012), Unsworth (2016), and Unsworth et al. (2018), we use experiential-based measures to establish language dominance for our participants and calculate this based on the information obtained by the PaBiQ as outlined in “The LITMUS-Questionnaire for Parents of Bilingual Children”.

### The LITMUS-Questionnaire for Parents of Bilingual Children

Bilingual children vary considerably in properties of their language exposure and use, which in turn influence the rate and outcome of their language development (e.g., Gathercole and Thomas, 2009; Chondrogianni and Marinis, 2011; Paradis, 2011; Hoff et al., 2012). Thus, having a clear idea about the relative amount of exposure and use for each of the bilingual child’s languages should help professionals to interpret language performance in L1 and L2 adequately and determine whether a child’s (poor) language performance is linked to possible risk factors for LI or to factors related to bilingualism such as the



timing, quality and quantity of exposure to the L1/L2, and degree of language dominance.

In order to gather relevant background information, the Questionnaire for Parents of Bilingual Children (PaBiQ; Tuller, 2015) developed during COST Action IS0804 on the basis of the Alberta Language and Development Questionnaire (ALDeQ, Paradis et al., 2010) and the Alberta Language Environment Questionnaire (ALEQ, Paradis, 2011) was used to interview the parents/legal guardians of the participating children. The parents of participants in the study were interviewed orally in their language of preference by trained native bilingual interviewers familiar with the respective culture.

The PaBiQ incorporates questions about developmental risk factors for SLI, which are synthesized into a global No Risk Index, for which a maximum of 23 points can be attained. This index is arrived at by collapsing the scores of the Positive Early Development index, which is associated with the timing of early language developmental milestones, and the Family History index, which is associated with the presence of oral/written language disabilities in the family. The Positive Early Development index (/14 pts) is calculated by adding up the sub-scores for age of first word ( $\leq 15$  mo = 6 pts; 16–24 mo = 4 pts;  $> 25$  mo = 0 pts), age of first multiword utterances ( $\leq 24$  mo = 6 pts; 25–30 mo = 4 pts;  $> 31$  mo = 0 pts) and early parental concerns (yes = 0 points; no = 2 points). The familiar risk for SLI (/9 pts) is indexed by the existence of first-degree relatives (mother, father, siblings) with reading/writing problems, difficulties understanding others when they speak or difficulties expressing themselves orally. Children with a negative family history of language problems are awarded a maximum of 9 points ( $3 \times 3$ : 1 point per family member per type of language difficulty). Boerma and Blom (2017) investigated the influence of LI and bilingualism on the latter two indices and looked into their diagnostic accuracy. In line with Paradis et al. (2010), they reported strong negative effects of LI on Early Language Development and showed that it was a strong predictor of LI in both monolingual and bilingual children confirming previous findings that a late onset of first words and sentences in at least one language is a risk factor for SLI (cf. de Houwer, 2009; Reilly et al., 2010). With regard to the Family History index, Boerma and Blom (2017) observed a negative effect for LI in the monolingual group but not in the bilingual one and concluded that, due to cultural factors, “Family History as reported by parents may [...] be less reliable as an index of LI in bilingual children than in monolingual children” (p. 73). The Positive Early Development Index also yielded promising diagnostic results in the study by Tuller et al. (2018), who found it be the leading factor explaining performance differences between BiSLI and BiTD in both of the German and French LITMUS non-word and sentence repetition tasks.

The PaBiQ further allows the calculation of a Language Dominance Index (LDI) as a differential between the L1 Exposure Index (relative amount of exposure to the L1) and the L2 Exposure Index (relative amount of exposure to the L2, i.e., German). For each of the child’s languages a total of 50

exposure/use<sup>4</sup> points could be attained using the German PaBiQ<sup>5</sup>. The Exposure Index is calculated for each of the child’s languages separately based on AoO, LoE<sup>6</sup>, frequency of early language use and exposure<sup>7</sup>, i.e., before the age of four, language richness before the age of four as measured by diversity of language exchange contexts, current language exposure/use within the family, current language use/exposure during different activities within an average week and in exchanges with playmates and family friends. The latter composite score also counts as an estimate of current language richness. An Exposure Index (/50 points) for L1 and L2 emerges by adding up the aforementioned sub-scores. A visual representation of the relative contribution of each of the sub-scores toward establishing the Exposure Index is given in **Figure 1**. As can be seen in **Figure 1**, current language use/exposure contributes the lion share (60%) to the calculation of the Exposure Index and consequently the LDI. This converges with the findings of Bedore et al. (2012) in their large-scale study, in which estimates of current language use (a composite score based on children’s amount of exposure and language output) accounted for 60% of the variance in language dominance patterns of bilingual children.

The language dominance index is then obtained by subtracting the L1 Exposure Index from that of the L2 yielding an estimate of the child’s degree of L2-dominance on a scale from –50 (extremely dominant in the L1) to +50 (extremely dominant in the L2). De Almeida et al. (2017, p. 5) compared multiple LDI cut-offs around LDI = 0 (optimal balanced bilingual) against impressions of bilingual investigators of the individual children after interacting with them and their families in both of their languages, and defined cut-off points for language dominance in attempt to explore the use of this variable. An LDI between –5 and +5 was set as a cut-off separating dominant from balanced bilinguals. Children with LDIs ranging from –5 to +5 are classified as “balanced,” children whose LDI is below –5 are considered to be dominant in the home language, while children with an LDI above +5 are classified as dominant in the societal language German.

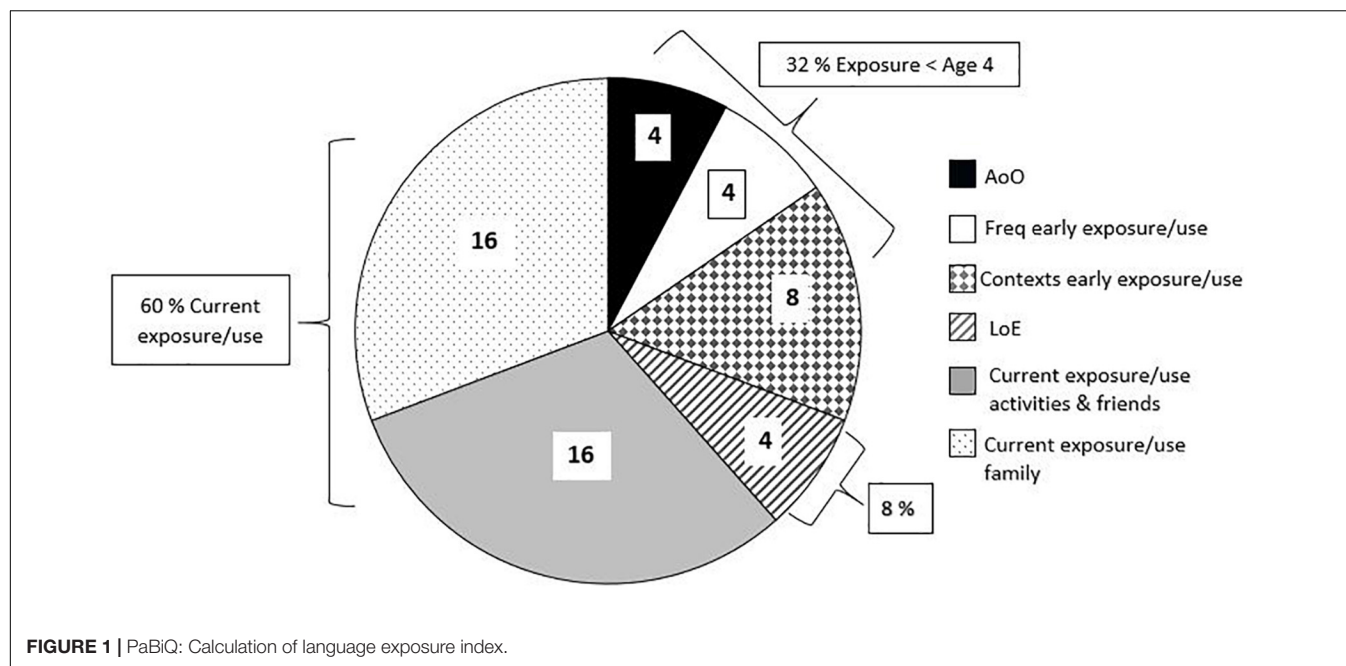
The questionnaire further allows determining the family’s socio-economic status (SES) based on the mother’s and the father’s educational levels. For the purposes of the current paper, maternal rather than paternal educational level (as measured by years of education of the mother) is used as a metric for SES, since the former is reported to be a strong predictor of language development, especially for expressive vocabulary levels, in both monolinguals (Hoff, 2003, 2006) as well as child bilinguals (Paradis, 2009; Calvo and Bialystok, 2014; Paradis and Jia, 2016; Meir and Armon-Lotem, 2017). SES-related language

<sup>4</sup>Input and output are collapsed together (interaction from interlocutor to child and from child to interlocutor).

<sup>5</sup>Five additional points were allotted to number of years in elementary school as part of the exposure indices in France, but not in Germany, where children join elementary school between the age of 6 and 8.

<sup>6</sup>Total length of exposure (LoE) is calculated by subtracting age of onset of systematic sustained exposure to the respective language from the chronological age.

<sup>7</sup>Contrary to Unsworth (2016) and Unsworth et al. (2018), PaBiQ’s language use estimates were not only limited to the “inside home context” but also cover the “outside home context.”



deficits<sup>8</sup> are reported to have a negative effect on performance in tasks with rich linguistic load, e.g., SRTs and NWRs with word-like items (Roy et al., 2014; Chiat and Polišenská, 2016).

## The German LITMUS Sentence Repetition and Non-word Repetition Tasks

### The German LITMUS Sentence Repetition Task

The German LITMUS-SRT (Hamann et al., 2013) used in this study was constructed in close parallel to the French LITMUS-SRT (de Almeida et al., 2017; Fleckstein et al., 2018). It consists of 45 sentences divided in three levels of syntactic complexity (five conditions per level controlled for syllable number, three test items per condition). The degree of an item's structural complexity relies on the presence of syntactic operations such as Wh-movement, clausal embedding, intervention<sup>9</sup> – where the latter may add difficulty to the presence of two propositions. Accordingly, level 1 consists of simple declaratives (7–9 syllables) and focuses on Subject-Verb-Agreement (SVA), tense and the sentence bracket[see (1)]. Level 2 (9–13 syllables) includes two types of object questions: bare Wh-questions with the non-D-linked *wh*-operator (*Wen* “who-masc.-acc.”), and *Welchen* NP-questions with the discourse-linked *wh*-operator (*Welchen* “which-masc.-acc.”) followed by an intervening lexical noun phrase [see (2a) & (2b)]. Bare Wh-questions are considered to be structurally less complex since

they do not involve intervention. Level 2 further contains non-finite and finite [see (3)] complement clauses. The latter are contrasted with coordinate structures, which serve as control items (two propositions but no embedding). Level 3 (11–12 syllables) comprises the most complex constructions and tests long passives, topicalizations [see (4)] as manifestations of the V2-property<sup>10</sup> of German, subject relative clauses as well as object relative clauses with [see (5)] and without intervening lexical determiner phrases.

Note that German has morphological case marking on accusative masculine singular pronouns, such as the interrogative and relative pronouns in examples 2a, 2b, and 5. **Table 1** gives an overview of test conditions. For more details on German

<sup>10</sup>German is a verb-second (V2) language, which requires the finite verb (marked for tense and agreement) to be the second constituent of main (root) clauses. Like other Germanic languages (except for English), it allows different types of constituents (e.g., complements, adverbials) to occupy sentence initial position in main clauses (e.g., topicalization), where only one constituent can linearly precede the finite verb (cf. Chomsky, 1986; Eisenberg, 1999).

**TABLE 1 |** German LITMUS-SRT: Overview of test conditions.

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.)

<sup>8</sup>A number of studies such as Balladares et al. (2016) showed that the influence of SES on repetition tasks diminishes after controlling for vocabulary sizes indicating that the effect of SES is primarily ascribed to smaller vocabulary sizes in children with lower SES.

<sup>9</sup>Intervention can be defined syntactically as in Rizzi (2004, 2013) modeled on work by Gibson (1998) or Gordon et al. (2001).

LITMUS-SRT, we refer to Hamann et al. (2017) and Hamann and Abed Ibrahim (2017).

(1) *Sentence bracket:*

Der Prinz **hat** die Prinzessin **umarmt**  
The/nom. prince has the/acc. princess hugged  
“The prince hugged the princess”

(2a) Bare WH

**Wen** beißt der große Löwe immer?  
Who/acc. bites the/nom. big lion always?  
“Who(m) does the big lion always bite?”

(2b) Which-NP

**Welchen Bauern** ärgert der Affe?  
Which/acc. peasant annoys the/nom. monkey?  
“Which peasant does the monkey annoy?”

(3) Finite complement clause:

Der Wikinger glaubt, **dass** die hexe ihn **mag**.  
The/nom. viking believes, that the/nom. witch him likes  
“The viking believes that the witch likes him”

(4) Topicalization

**Den Arzt** fotografiert der Bauer gerne  
The/acc. doctor photographs the/nom. peasant gladly  
“The doctor, the peasant photographs gladly”

(5) Object relative with intervention:

Ich sehe den Vogel, **den der Pinguin** weckt.  
I see the/acc. bird who/acc. the/nom. penguin wakes up  
“I see the bird who(m) the penguin wakes up”

The test stimuli are pre-recorded, pseudo-randomized and integrated into a child friendly PowerPoint Presentation. The administration of the task takes about 10 minutes. The task is scored both by identical repetition of test items (SRT\_Id), i.e., whole item accuracy, where only phonological errors are disregarded, and by correct target structure (SRT\_Tar), which measures whether a particular structure has been mastered or not (see Marinis and Armon-Lotem, 2015 for scoring measures). Although scoring by SRT\_Id is faster and easier, L2-errors not affecting the realization of the targeted structure such as lexical substitutions, omissions and systematic recurrent case<sup>11</sup> as well as gender errors could surface using this scoring method and penalize bilingual children. Comparison of these scoring methods has indeed shown that SRT\_Tar leads to higher diagnostic accuracy of the test for German (see Hamann and Abed Ibrahim, 2017 for particulars).

### The German LITMUS Non-word Repetition Task

The German LITMUS-NWRT (Grimm et al., 2014) employed in this study is composed of two parts: a structurally less complex (quasi-) language independent part (NWRT\_LI) and

a language dependent part (NWRT\_LD) incorporating more complex structural aspects. In both parts the item length ranges from one to three syllables with constant word-initial stress. The 30 items of the LI part were constructed using phonemes and phonotactic constraints attested in the vast majority of the world's languages (Maddieson et al., 2011), i.e., phonemes that are “compatible with cross-linguistically diverse constraints on lexical phonology” (Chiat, 2015, p. 138). Unlike the non-words of the Quasi-Universal-NWRT discussed in Chiat and Polišenská (2016), the non-words of the German LITMUS-NWRT are shorter and are not only composed of simple CV sequences, but also include syllables with initial consonant clusters “#CCV” or closed syllables of the type “CVC#,” which are typologically well-attested albeit their relative complexity (Maddieson, 2006). Throughout the task, phonological complexity is systematically varied at the segmental (consonantal), syllabic (presence of branching onsets or coda) or sequential (position of cluster within the non-word) levels (see dos Santos and Ferré, 2018; Grimm and Hübner, in press for details). The LD part contains 36 items adhering to the same construction principles of the LI part in addition to the extrametrical /s, ʃ/ in word initial and final positions as a complexity feature specific to German (and some other languages, e.g., English and Russian). Such sC sequences violate the Sonority Sequencing Principle and are considered phonologically more complex than other types of onset clusters. Constructed as such, the LD part is considered to be structurally more complex compared to the LI part, yet less dependent on LS knowledge than the more traditional Language-Specific NWRTs, e.g., Rispens and Baker (2012), which draw on the full phoneme inventory (consonants and vowels) and include many more properties specific to the target language (Chiat, 2015; Chiat and Polišenská, 2016).

Although structures with higher phonological complexity are generally more error-prone in TD children, they are “disproportionately difficult” for children with SLI (Chiat, 2015, p. 137), who struggle with phonological complexity (Archibald and Gathercole, 2006; Jones et al., 2010; dos Santos and Ferré, 2018). Thus, a greater performance gap between TD and SLI is expected for both monolingual and bilingual children on NWRT\_LD, which contains trilateral sCC onset clusters, where /s/ and /ʃ/ represent an appendix to the prosodic word. The latter has been shown to be deficient in phonologically impaired monolingual German children (Ott et al., 2006). An overview of segments and syllable types is given in Table 2.

Task administration takes about 5 min and the non-words are presented to the child in a pseudo-randomized order via an animated PowerPoint Presentation. At the beginning of the task, children are provided with noise-canceling headphones and are told that an alien from another planet would appear on the screen and try to teach them his language (format adapted from Engel de Abreu et al., 2013). The test is scored by whole item accuracy (percentage of items correct), since this scoring method is better suited for clinical purposes and has been shown to be informative (Roy and Chiat, 2004; Boerma et al., 2015). A response is rated as correct if all consonants and vowels in addition to their sequencing correspond to the target form. Phoneme omissions, substitutions or additions are regarded as incorrect. Systematic

<sup>11</sup>Case errors are not disregarded if they are crucial for the realization of the targeted structure, e. g., object relatives and topicalized sentences.



**TABLE 2 |** Overview of segments and syllable types in German LITMUS-NWRT.

	Vowels	Consonants	Syllable types	Examples
Language-Independent part (LI)	/a, i, u/	/p, k, f, l/	CV CCV CVC#	kapi plaku pukif
30 items				
23 test items				
7 controls (e.g., faku, paf)				
Language-Dependent part (LD)	/a, i, u/	/p, k, f, l/	same syllable types plus	
36 items		plus	#sCV	sfikupla
32 test items		/s/	#sCCV	sklipafu
4 controls (e.g., kif, sapi)		/ʃ/	Cs# internal /s/	kapifaps fikuspa

phoneme replacements reflecting articulatory difficulties, e.g., /t/ for /k/ (/kafip/→/tafip/) are not counted as errors. Since the task mainly targets bilingual children, L2-errors such as voicing of consonants (/pilu/→/bilu/) or vowel alternations (/faku/→/fako/) are disregarded. Furthermore, substitution of extrametrical /ʃ/ through [s] or an interdental pronunciation of extrametrical /s/ are not counted as errors since this does not result in a phonemic contrast in extrametrical positions in German (Grimm and Hübner, *in press*).

## Participants

The present study was conducted in line 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). Parents or legal guardians of all participating minors provided written informed consent for both data collection and analysis. The research protocol was approved by the “Kommission für Forschungsfolgenabschätzung und Ethik” of the Carl-von-Ossietzky University of Oldenburg.

Except for 3 children, the current study used the same participant sample as Hamann and Abed Ibrahim (2017), including 77 children, 21 German monolinguals and 56 L2-German bilinguals with Arabic, European Portuguese or Turkish as L1. The latter L1s were chosen because a sizable proportion of immigrants residing in Germany are of Arab, Portuguese and Turkish origin. Furthermore, the typological differences between them and the children’s L2 (German) enable cross-group comparisons, e.g., Abed Ibrahim et al. (2018) and Chilla et al. (*in press*). The age range of the participants was 5;6–9;0 years covering the last year of kindergarten and the crucial first 2–3 years of primary school. As inclusion criteria for bilingual children, children had to have a minimum L2 exposure of 18 months and be at least functionally bilingual. Thus, children who failed to complete even receptive subtests in the L1 were excluded from the study. 49/56 children were simultaneous bilinguals, while 7 were sequential bilinguals, whose systematic

exposure to L2 mainly started upon kindergarten entry at approximately age three. Almost all of the bilingual participants had a LoE to German of more than 24 months at the time of testing with a mean LoE of 5;1 years ( $SD = 1;10$ ). Children likely to have SLI, i.e., with a clinical diagnosis of SLI, were recruited from specialized speech-language pathology centers and kindergartens with special inclusion programs from different parts of Germany. Given the high rates of over- and under-referral of bilingual children to speech language therapy (Grimm and Schulz, 2014), an extensive procedure based on standardized evaluation in each of the child’s languages was applied in order to verify the clinical status of all recruited bilingual children as  $\pm$  language-impaired. The verification of clinical status was done in accordance with the recommendations of the COST Action IS0804 assessment committee as outlined in Thordardottir (2015, p. 343) and began with a control for non-verbal intelligence using the German version of Raven’s Colored Progressive Matrices (CPM; Bulheller and Häcker, 2002). Only Children who had a non-verbal IQ score  $\geq 80$  were included in the study. In addition to standardized assessment, narrative samples were collected from each child in both of her languages using the picture materials provided by the LITMUS-Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2015). The collection of the narrative samples was done in accordance with the MAIN protocol (story telling). However, for the purposes of the current study, the latter samples were not analyzed in terms of narrative macro- and microstructure, but were rather used as spontaneous speech samples. Especially in borderline cases, the latter samples were consulted in order to gain an impression<sup>12</sup> about the child’s expressive language abilities in both of her languages and look for clinical markers for SLI, e.g., SVA errors, the use of infinitives and verb placement errors in German (Clahsen, 1991; Rice et al., 1997; Hamann et al., 1998; Lindener, 2002).

As to assessment using formal tests, in our previous work, e.g., Hamann and Abed Ibrahim (2017), Tuller et al. (2018), and Chilla et al. (*in press*), we adapted the criteria outlined in Leonard (2014) to bilinguals using Thordardottir’s (2015) recommendations and assigned a child to the BiSLI group if she scored below dominance-adjusted<sup>13</sup> norms in two language domains (on norm-referenced tests) in both of her L1 and L2. Five language areas relevant in this context were evaluated in each of the child’s languages (except for Turkish): phonology, morphosyntax comprehension and production as well as receptive and expressive vocabulary (see also Tomblin et al., 1996). Since expressive vocabulary is a notorious locus of difficulty for bilingual children, we counted lexicon as a single domain and considered the child unimpaired in this domain

<sup>12</sup>The narrative samples of the children were evaluated by linguistically trained native speakers (L1 and L2) according to certain markers: e.g., subject-verb-agreement and sentential complexity, i.e., presence of embeddings.

<sup>13</sup>Following the recommendation of Thordardottir (2012, 2015), the monolingual  $-1.25$  SD cut-off criterion used by Tomblin et al. (1996) was adapted according to the dominance status of the language being assessed. Accordingly, we used a criterion of  $-1.5$  SD if the child was evaluated in her dominant language, and a cut-off of  $-2.25$  SD if the child’s weaker language was being assessed. In case of balanced bilinguals, the cut-off criterion was set at  $-1.75$  SD for both languages.



**TABLE 3 |** Overview of norm-referenced tests employed for standardized language assessment in Arabic, German, European Portuguese, and Turkish.

Language	Test	Language skill evaluated					Method of scoring	Age range
		Phonology	Reception vocabulary	Expression 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	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 Bilingual: 3;0–7;11
	PLAKSS-II <sup>d</sup>	Picture naming	–	–	–	–	Individual subtest scores	2;6–7;11
European Port.	PALA-P <sup>e</sup>	Non-word repetition	Picture selection	Picture naming	Picture selection	Sentence repetition	Individual subtest scores	5;0–9;11 (with missing norms for some age ranges for all tasks)
	GOL-E <sup>f</sup>	–	Word definition	Antonyms naming	–	Complex S from two simple S's	Individual subtest scores and global score	5;07–10;00
Turkish	TEDIL <sup>g</sup>	–	Picture selection	Picture naming	Picture Selection	Sentence completion/construction	2 composite scores, 1 production and 1 comprehension	2;0–7;11

<sup>a</sup>Zebib et al. (2017); <sup>b</sup>Glück (2011); <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); and <sup>g</sup>Topbaş and Güven (2013); TVJT, truth value judgment task.

if only receptive vocabulary was above the respective cut-off. For the assessment of L1 and L2, we chose norm-referenced L1 and L2 tests frequently used by speech language pathologists and cover the age range<sup>14</sup> under investigation (see **Table 3** for a detailed overview of standardized assessment tools). For German, we selected the LiSe-DaZ (Schulz and Tracy, 2011), which provides bilingual and monolingual norms, for assessing morphosyntax. The short form of the WWT (Glück, 2011) was used to assess receptive and expressive vocabulary, and the screening version of the PLAKSS-II (Fox-Boyer, 2014) was used to evaluate phonology. We tried to assess the same language domains in Arabic, Portuguese and Turkish. For Arabic, this was possible using the comprehensive test battery ELO-L (Zebib et al., 2017), which offers norms for Lebanese Arabic and was adapted to a number of other varieties of Arabic<sup>15</sup> by the test authors in collaboration with linguistically trained native speakers of the respective varieties (Algerian, Iraqi, Libyan, Moroccan, Palestinian, Syrian, and Tunisian). We used the PALPA-P test battery (Castro et al., 2007) for Portuguese. One major limitation

of the PALPA-P is that it lacks norms for some of the age ranges we are investigating for the lexical domain. As a result, we chose to assess receptive and expressive vocabulary using subtests of the GOL-E (Sua-Kay and Santos, 2014), which covers our entire age range, and used subtests of the PALPA-P to assess phonology and morphosyntax. For Turkish, we chose the TEDIL (Topbaş and Güven, 2013), which measures morphosyntactic comprehension and production as well as lexical semantics. The test; however, does not include a subtest for phonology and does not offer norms for the individual subdomains. Instead, a composite score exists for each of comprehension and production collapsing morphosyntax and lexical semantics together. As the Turkish test merely offers a single production and a single comprehension score, encompassing two domains each, a child was assigned to the BiSLI group if she scored below cut-off in either production or comprehension. For a detailed description of standardized assessment L1-L2-tests and a complete overview of recruitment and classification procedure of bilingual children into TD vs. SLI, we refer to Hamann and Abed Ibrahim (2017).

Following the argumentation in Hamann and Abed Ibrahim (2017, p. 16) about problems encountered with standardized L1 tests in heritage contexts, and since our previous classification procedure did not isolate subgroups of SLI and might have missed cases of selective impairment such as grammatical/syntactic, phonological or lexical SLI (cf. Friedmann and Novogrodsky, 2008), we adopted Hamann and Abed Ibrahim's (2017) modified

<sup>14</sup>In case of children older than the norming sample, we consulted with the test authors concerning the possibility of norm-extension, e.g., LiSe-DaZ and ELO-L.

<sup>15</sup>Norms are only available for the Lebanese version. Due to linguistic proximity between Lebanese, Syrian, Jordanian and Palestinian Arabic, the Lebanese norms can be applied to the latter varieties with the caveat that the socio-cultural context may differ. In case of Maghreb (Moroccan, Tunisian, and Algerian) dialects, norms should be viewed with caution, especially in borderline cases.

**TABLE 4 |** Participants including monolingual controls and bilinguals after verification of clinical status (Mean, SD and range<sup>16</sup>).

Simult./total	MoTD (n = 10)	MoSLI (n = 11)	BiTD (n = 44)				BiSLI (n = 12)
			BiTD-A (n = 10)	BiTD-P (n = 18)	BiTD-T (n = 16)	Total (n = 44) 38/44	
Age at testing	75.90	79.00	88.60	83.44	86.68	85.6	80.66
(in months)	(8.99)	(9.89)	(13.5)	(14.72)	(13.01)	(13.60)	(15.05)
	66–92	68–98	70–108	66–108	70–104	66–108	64–108
Age of onset			39	17.61	25.31	25.27	16.00
(in months)	0	0	(13.61)	(24.39)	(15.52)	(20.65)	(17.03)
			24–75	0–90	0–48	0–90	0–36
Length of exposure	75.90	79.00	53.2	67.97	61.37	61.73	62.00
(Gr.) (in months)	(8.99)	(9.89)	(19.68)	(24.59)	(19.01)	(21.87)	(22.77)
	66–92	68–98	32–97	18–101	34–96	18–101	30–88
CPM (PR)	81.20	53.72	44	73.77	71.18	66.06	55.08
	(13.98)	(24.50)	(29.13)	(17.69)	(22.23)	(25.01)	(26.69)
	56–100	25–99	9–93	42–94	38–100	9–100	27–98
LDI (/50)			–7	1.44	–5.75	–3.09	0.58
	N/A	N/A	(9.34)	(13.76)	(12.57)	(12.77)	(12.37)
			–25–11	–25–23	–27–18	–27–23	–21–24
L1-dominant (no./total)	N/A	N/A	(5/10)	(8/18)	(8/16)	(21/44)	(3/12)
Balanced (no./total)	N/A	N/A	(1/10)	(3/18)	(4/16)	(8/44)	(5/12)
L2-dominant (no./total)	N/A	N/A	(4/10)	(7/18)	(4/16)	(15/44)	(4/12)
Yrs. educ. mother			14.2	13	13.68	13.52	11.83
	N/A	N/A	(1.68)	(3.91)	(5.23)	(4.05)	(2.75)
			12–18	4–18	5–20	4–20	8–16

<sup>16</sup>When applicable.

“criteria for the identification of the bilingual clinical group” in this paper. Accordingly, we assigned a child to the BiSLI group if she had a selective impairment in the L2, i.e., if she performed below the dominance-adjusted cut-off in either morphosyntax or receptive vocabulary or phonology (not necessarily two domains in combination), and scored below norms in two domains in her L1 (one domain for Turkish) or showed poor performance of spontaneous production in both of her L1 and L2. **Table 4** gives a participant overview based on clinical status as verified by the modified procedure described above and also includes the two monolingual control groups MoSLI and monolingual typically developing children (MoTD). By applying the modified classification criteria, the clinical status of 4 children who were initially classified as BiTD in Hamann and Abed Ibrahim (2017) changed to BiSLI<sup>16</sup>. In **Table 4**, the BiTD children are divided into subgroups based on their L1: Arabic = BiTD-A, Portuguese = BiTD-P and Turkish = BiTD-T. The BiSLI group is composed of 12 children (4 with L1 Arabic, 3 with L1 Portuguese and 5 with L1 Turkish). Due to the relatively small sample size, the BiSLI children are grouped together regardless of their home language. The bilingual children are further classified according to language dominance as measured by the PaBiQ (see section

“The LITMUS-Questionnaire for Parents of Bilingual Children”). As can be seen in **Table 4**, almost half of the children in the BiTD group (21/44) are dominant in their L1, whereas the majority of the BiSLI children (9/12) are either balanced or L2-dominant<sup>17</sup>.

The four groups (MoTD, MoSLI, BiTD, and BiSLI) were comparable in terms of non-language variables such as chronological age and non-verbal intelligence. Concerning age, the overall effect of Group was not significant, as revealed by Kruskal–Wallis test [ $\chi^2(3, N = 77) = 5.505, p = 0.138, \eta^2 = 0.034$ ]. This also holds when the BiTD group is split into three subgroups by L1 [ $\chi^2(5, N = 77) = 6.758, p = 0.239, \eta^2 = 0.051$ ]. In terms on non-verbal intelligence, the overall effect of Group was significant [ $\chi^2(3, N = 77) = 8.448, p = 0.038, \eta^2 = 0.075$ ]. However, subsequent pairwise comparisons using Mann–Whitney U tests controlling for false positives, that is Type I error, revealed only one marginally significant comparison, namely MoSLI vs. MoTD ( $U = 19.00, p = 0.06, r = 0.553$ , Bonferroni-corrected). Yet, all of the children belonging to the MoSLI group have normal non-verbal intelligence. We further checked whether the bilingual groups were comparable concerning SES, AoO, LoE, and degree of L2-dominance (LDI). No significant differences emerged between BiTD and BiSLI concerning SES [ $\chi^2(1, N = 56) = 2.228, p = 0.135, \eta^2 = 0.041$ ], AoO [ $\chi^2(1, N = 56) = 3.261, p = 0.071$ ,

<sup>16</sup>Two of them were cases of selective impairment in the L2, i.e., they performed below cut-off in the L1, and only showed deficits in morphosyntax in the L2 (grammatical SLI) plus SLI markers in the speech samples in both languages. The other two were L2-dominant and performed below-cut-off in all of the domains in the L2, but slightly above cut-off in the L1. Since SLI markers were present in their speech samples, we classified them as BiSLI

<sup>17</sup>This reflects the advice frequently given to parents of bilingual children with atypical language development that they should restrict parent-child interactions to the societal language to avoid aggravating the existing language difficulties, which in turn means less exposure to the L1.

$\eta^2 = 0.059$ ], LoE [ $\chi^2(1, N = 56) = 0.615, p = 0.433, \eta^2 = 0.011$ ], and LDI [ $\chi^2(1, N = 56) = 1.912, p = 0.167, \eta^2 = 0.035$ ]. This also holds when the BiTDs are split by L1 SES [ $\chi^2(3, N = 56) = 3.216, p = 0.360, \eta^2 = 0.06$ ], LoE [ $\chi^2(3, N = 56) = 3.640, p = 0.303, \eta^2 = 0.07$ ] and LDI [ $\chi^2(3, N = 56) = 4.457, p = 0.216, \eta^2 = 0.08$ ]. With respect to AoO, the overall effect of Group was significant when BiTDs were divided by L1 into three subgroups L1 [ $\chi^2(3, N = 56) = 11.833, p = 0.008, \eta^2 = 0.17$ ]. Mann-Whitney U tests applying Bonferroni-adjustment of  $p$ -values revealed significant differences in AoO between BiTD-A and BiTD-P ( $U = 33.00, p < 0.05, r = 0.531$ ) as well as between BiTD-A and BiSLI ( $U = 17.00, p < 0.05, r = 0.617$ ). Nevertheless, the overall effect of Group was not significant when the BiTD groups were collapsed together [ $\chi^2(1, N = 56) = 3.261, p = 0.071, \eta^2 = 0.059$ ].

## Data Analysis

The children's responses on the SRT and NWRT were recorded using special dictaphones. Data transcription, verification and coding for errors were done offline by two independent linguistically trained raters (percentage of agreement was at least 90%). For each repetition measure, the percentage of correct responses was used as basis for data analysis. Null reactions were counted as errors, unless they were due to technical problems or errors by the investigators (missing data, less than 1% of the overall data).

IBM SPSS 24 (2016) and R-Studio (2012) were used to conduct statistical analyses. Non-parametric tests were used for group comparisons due to unequal sample sizes and the violation of the normality assumption, checked by the Shapiro-Wilk test. Since we wanted to investigate whether the LITMUS repetition tools are suitable for assessment of bilingual children in their weaker language, we first checked for group differences between L1-dominant BiTDs and their monolingual, balanced and L2-dominant TD peers, and whether performance of L1-dominant BiTDs overlapped with that of MoSLIs and BiSLIs. Here, we split the BiTDs into three subgroups based on LDI as established in the section "The LITMUS-Questionnaire for Parents of Bilingual Children",<sup>18</sup> and ran Kruskal-Wallis tests and Mann-Whitney U tests with Bonferroni-adjustment. Recall that BiSLIs were collapsed into a single group due to the small sample size. Since performance of BiTDs on SRT appeared to be influenced by dominance, we ran partial correlation analysis controlling for age on their SRT\_Id and SRT\_Tar. In addition to language dominance, we also checked for correlations with AoO, LoE and SES, since they are factors known to influence performance on linguistic tasks. Next, linear regression models for predicting performance of the BiTDs on SRT\_Id and SRT\_Tar were built using the variables that yielded significant correlations.

Secondly, we applied cluster analysis to the data in order to automatically group the children into  $\pm$  language-impaired based on their performance scores on the SRT (SRT\_Id, SRT\_Tar) and NWRT (NWRT\_global, NWRT\_LI, NWRT\_LD), separately and then in combination. A clustering algorithm classifies a dataset into several meaningful homogenous sub-categories - so-called clusters (i.e., TD vs. SLI in this study) - based on the values of their

attributes (i.e., linguistic variables in the present study) such that the similarity<sup>19</sup> among objects within a category is larger than that between categories. We opted for unsupervised learning (cluster analysis) for verifying diagnostic accuracy and establishing cut-off points separating TD from SLI on the tasks, since it does not use predefined clinical status during the statistical analysis, and is thus unbiased by any given classification of participants.

Because children were measured based on performance scores on LITMUS-SRT and NWRT designed to identify SLI without penalizing bilinguals, our premise was that SLI-cases would be similar to each other, and hence group together, while TD-cases would form their own cluster regardless of bilingualism. Different from Hamann and Abed Ibrahim (2017), we chose the PAM (Partitioning Around Medoids) non-hierarchical k-medoid clustering method (Kaufman and Rousseeuw, 1987, 2009) over k-means, because it is a suitable method for small datasets with up to approximately 60 objects, and because it can handle noisy data and outliers (Kaufman and Rousseeuw, 1987, 2009; Kashef and Kamel, 2008; Patel and Singh, 2013; Soni and Patel, 2017). Variables were scaled for normalization purposes in the course of the PAM-analysis. We used the function *pam* of the *cluster* R package (Maechler et al., 2017).

We used *Hopkins statistic* (H) based on the *factoextra* R package (Kassambara and Mundt, 2017) as a measure of cluster tendency to assess clusterability (Hopkins and Skellam, 1954). If the H-value is close to zero, and far below 0.5, then the dataset is clusterable (Kassambara and Mundt, 2017; Krishna et al., 2018). Because H is run on the created random dataset every time, we get fluctuations in the H-values if we run the statistics multiple times. Banerjee and Davé (2004) demonstrate that random data sets, clustered data sets and regularly spaced data sets show H-values of around 0.5, 0.7–0.99 and 0.01–0.3, respectively.

Because the k-medoid algorithm requires that the number of clusters should be pre-defined, we first ran the *Gap Statistic* (Tibshirani et al., 2001) to determine the optimal number of clusters. The *Gap Statistic* compares the change in within-cluster dispersion for each clustering solution (at each number of clusters) to that expected at random distribution. We used the functions *fviz\_nbclust* of the *Factoextra* R package (Kassambara and Mundt, 2017) and *NbClust* of the *NbClust* R package (Charrad et al., 2014) to determine the optimal number of clusters.

The k-medoid algorithm selects one of the members of the cluster as the most representative object, named *cluster medoid*, so that each cluster has only one medoid. By choosing an actual case (i.e., an SLI or a TD child) as the cluster medoid, the k-medoid method is less sensitive to outliers, as mentioned before. The optimal cluster is achieved by minimizing the sum of squared Euclidean distances to the medoid in each cluster, also called the error sum of squares (Kaufman and Rousseeuw, 1987). First, in the so-called "Build-step," the k-medoid algorithm selects k medoids randomly, with k being the optimal number of clusters. Next, a matrix of dissimilarity is calculated from the raw data and the algorithm assigns every object to either of the

<sup>18</sup>Note that language dominance was used as a categorical variable in this step.

<sup>19</sup>The notion of similarity in the clustering approach is operationalized as Euclidean distance.

k clusters based on their distance to the nearest medoid (Patel and Singh, 2013). The sum of absolute error in the clustering procedure is equal to the sum of the distances between data points and their medoids. In the so-called “Swap-step,” each non-medoid object is tested as a potential medoid within each cluster by checking if the sum of within-cluster distances gets smaller if that object is used as the new medoid. If this is the case, then that configuration is used. The algorithm checks at each iteration step, if the solution is better than the previous one. If the medoids do not change, the algorithm terminates (see Patel and Singh, 2013 for details).

Because the medoid of each cluster can be seen as a prototype of that cluster, identifying the medoid can serve as a cue to interpret the cluster. For example, if the medoid of a cluster was originally diagnosed as an SLI-case, then that cluster represents most probably the SLI-cases. We expected the SLI-cluster to contain the majority of the children classified originally as SLI based on standardized assessment, while the majority of TD-cases would reside in the other larger cluster. Our further premise was that the cluster with the lower scores on the linguistic variables would represent the cluster with LI, since language-impaired cases score lower on the linguistic variables.

After clustering the sample, we determined the estimated cut-offs on the linguistic variables (i.e., SRT and NWRT) between the SLI- and TD-clusters based on the clustering result. A cut-off is a value of a variable which can be seen as the best threshold score to separate the cases belonging to the two categories using that variable. If the two categories can be best separated along multiple variables simultaneously, e.g., SRT and NWRT combined, then cases can be predicted (as TD vs. SLI in our study) based on multiple cut-offs on these variables. To this end, we employed conditional inference tree models (Tagliamonte and Baayen, 2012). Conditional inference trees (ctrees) are non-parametric regression models visualized as decision trees. They are suitable for our dataset because of the presence of high-order interactions among the variables and the overall small sample-size compared to the number of predictors (Levshina, 2015). Besides determining the cut-off for the linguistic variables, ctrees can also give information about the hierarchical structure of the relevant predictors of cluster membership, i.e., about variable importance. For instance, if clustering is based on several linguistic variables such as SRT\_Id, SRT\_Tar and NWRT\_global, decision trees can show which one contributed the most toward predicting cluster membership as TD or SLI. The higher the variable in the hierarchy, the more important it is, with the highest-level variable being the most important. If there are multiple variables in the ctree, then a multi-hierarchy predicts the outcome (i.e., cluster membership as TD or SLI). Ctrees were implemented with the *party* R package using the *ctree* function (see Hothorn et al., 2006 for details).

In order to address research questions (ii) and (iii), we calculated diagnostic accuracy<sup>20</sup> for the SRT and NWRT measures separately and combined. Sensitivity and specificity levels were

estimated by comparing cluster membership of each of the children as TD vs. SLI as assigned by PAM on the basis of LITMUS-SRT and NWRT results to their clinical status (as established by the standardized assessment procedure described in “Participants”). Sensitivity is determined by the proportion of children with LI identified as such by LITMUS SRT and NWRT or subtests thereof (i.e., assigned to the clinical cluster in our case), while specificity is computed based on the proportion of children with typical language development identified as such by our tests, i.e., assigned to the non-clinical cluster (Oetting et al., 2008; Dollaghan and Horner, 2011). In addition, likelihood ratios<sup>21</sup> (LRs) were calculated based on the obtained sensitivity and specificity levels. An advantage of LRs is that they are less likely to be affected by variations in the properties of the test sample (Dollaghan and Horner, 2011). LR+, positive likelihood ratio [sensitivity/(1-specificity)], indicates how likely it is that a score below a cut-off criterion to be present in language-impaired children, whilst an LR–, negative likelihood ratio ((1-sensitivity)/specificity), is indicative of the likelihood of a score above a cut-off criterion to belong to a child without LI.

To answer research question (i), we investigated which of the background information variables provided by the PaBiQ as cogent confounders predicted cluster membership following each clustering procedure based on SRT and NWRT measures or combinations thereof. The hypothesis to be tested was that cluster membership as TD or SLI based on performance scores in LITMUS-NWRT and/or SRT can only be explained by variables concerning risk factors to SLI and not by background information variables related to bilingualism, particularly the degree of language dominance. If this hypothesis is confirmed, then the clustering of the cases cuts across the SLI/TD dimension rather than any of the background information variables unrelated to risk factors for SLI validating that the diagnostic accuracy of the tasks is not compromised by language dominance. To that end, we ran Firth’s Bias-Reduced Binary Logistic Regression (Firth, 1993), which uses penalized ML<sup>22</sup>. Cluster membership (TD or SLI) served as the dependent measure. Models with Firth’s correction were built using the *Brglm2* R package (Kosmidis, 2018). We included only a maximum of four background information variables as fixed factors in the model to avoid over-parametrization given the overall small sample size. Because regression analysis provides a way of adjusting for potentially confounding covariates included in the model, we entered the covariates into the model at once. To examine

<sup>21</sup>LR+ values  $\geq 10$  are highly indicative of the presence of language impairment, LR– values  $\leq 0.10$  highly indicate the absence of an impairment, LR+ values  $\geq 3.0$  and LR– values  $\leq 0.3$  are considered to be clinically suggestive, while LR+ values  $< 3.0$  and LR– values  $> 0.3$  are viewed as clinically uninformative (cf. Dollaghan and Horner, 2011).

<sup>22</sup>We used Firth’s Bias-Reduced Logistic Regression to step around the following statistical concerns. Given our small sample, Long (1997) advises against maximum likelihood estimation in logistic regression with less than 100 cases. A second confound is that in small samples maximum likelihood estimates in binary logistic regression models are not powerful because there can be a complete- or quasi-complete separation along one covariate (Rainey, 2016). That is, for a given combination of covariates the outcome can be predicted perfectly. To avoid separation, Peduzzi et al. (1996) suggest that the number of positive outcome events (i.e., the smaller number of binary outcomes) divided by the number of independent variables should be more than 10.

<sup>20</sup>Following Plante and Vance (1994), good diagnostic accuracy is given when specificity and/or specificity rates are  $\geq 90\%$ . Rates between 80 and 90% are viewed as fair.



whether the diagnostic accuracy of the tasks is not compromised by language dominance and is only sensitive to risk factors for SLI, we built several regression models using Firth's correction with PAM cluster membership as TD or SLI as the dependent variable. In each model, we entered LDI and the index of Positive\_Early\_Development (risk factors for SLI) in addition to two further background information variables reported to explain performance on LITMUS-SRT and NWRT (see Tuller et al., 2018) as covariates. The latter variables included AoO, LoE, SES. We also included chronological age as a covariate since working memory and cognitive capacities are rapidly growing in children and since language abilities of children tend to improve over time.

Background information variables were first scaled by the mean of their original variable to remove potential *non-essential* multi-collinearity between them (Dalal and Zickar, 2011) and to adjust the interpretation of the coefficients. Multi-collinearity among covariates was checked using the *Variance Inflation Factor (VIF)* after scaling, with a VIF value above 10 indicating serious multi-collinearity (Kutner et al., 2004). Correlations between the background variables are given in the **Appendix**.

## RESULTS

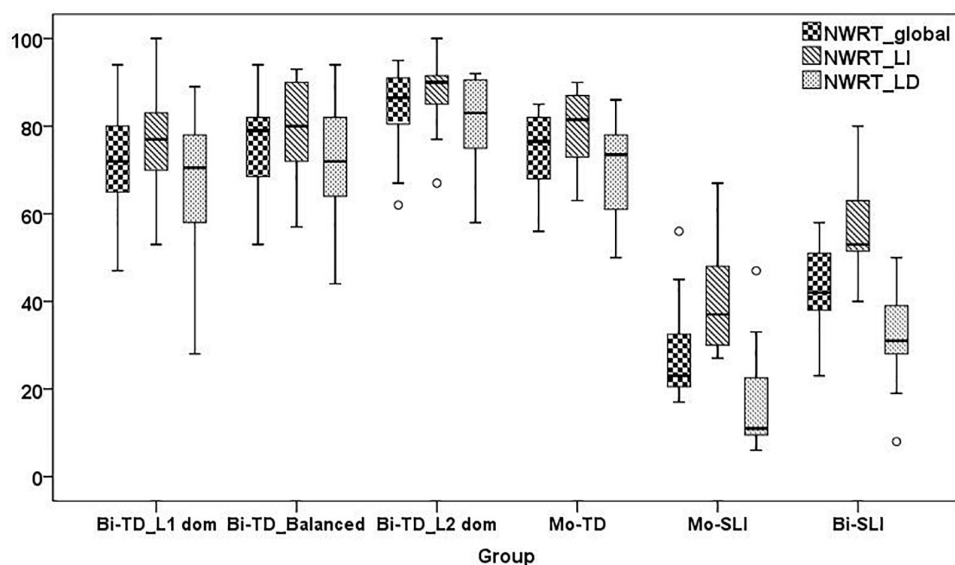
### Overall Results on the German LITMUS NWRT and SRT

Kruskal-Wallis tests comparing performance scores of L1-dominant BiTDs to the other groups (MoTD, balanced-BiTD, L2-dominant-BiTD, MoSLI, and BiSLI) on NWRT\_global, NWRT\_LI, NWRT\_LD, SRT\_Id and SRT\_Tar yielded significant results for all measures as shown in **Table 5**.

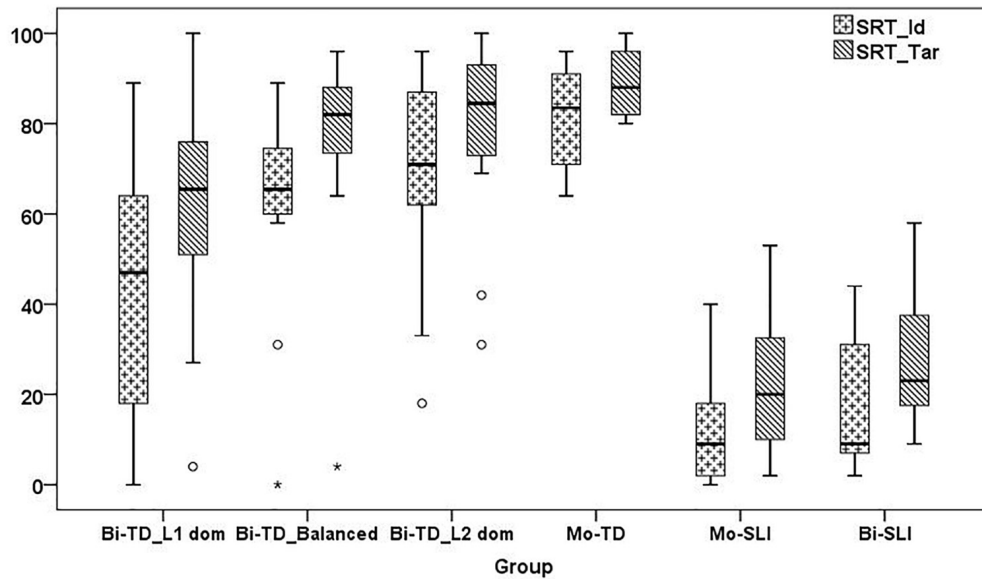
Subsequently, pairwise comparisons were carried out using Mann-Whitney U tests with Bonferroni-adjustment. Typically

developing children performed significantly better than their language-impaired counterparts on all measures. All measures distinguish between MoTDs and MoSLIs as well as between BiTDs and BiSLIs regardless of language dominance: Moreover, all of the BiTD groups significantly outperformed MoSLIs. The comparisons yielded no significant differences between MoSLIs and BiSLIs on any of the aforementioned measures. Comparing MoTDs to the BiTDs split by dominance revealed no significant differences between MoTDs and balanced as well as L2-dominant BiTDs on either measure. Nevertheless, significant differences with large effect sizes were found between MoTDs and L1-dominant BiTDs as well as between L1-dominant and L2-dominant BiTDs for both SRT\_Id and SRT\_Tar but not for any of the NWRT measures (see **Table 5**). It should, however, be stressed that despite the observed significant differences in SRT\_Id and SRT\_Tar, L1-dominant BiTDs performed significantly better than MoSLIs and BiSLIs on both SRT measures. **Figures 2 and 3** depict the overall performance of the groups in NWRT and SRT, respectively. An overview of significant pairwise comparisons is provided in **Table 5**.

In the next step, we collapsed all of the BiTDs into one group and ran partial correlation analysis controlling for age on their performance in SRT\_Id and SRT\_Tar and variables shown to influence performance on LITMUS-SRT including language dominance (see Tuller et al., 2018). Moderate positive correlations were found between LDI and performance on SRT\_Id ( $r = 0.542, p < 0.001$ ) and SRT\_Tar ( $r = 0.586, p < 0.001$ ), as well as SES and SRT\_Id ( $r = 0.478, p = 0.001$ ) and SRT\_Tar ( $r = 0.431, p = 0.004$ ). The analysis revealed a weak positive correlation between SRT\_Id and LoE ( $r = 0.364, p < 0.05$ ) and a weak negative correlation between SRT\_Id and AoO ( $r = -0.348, p < 0.05$ ), whereas the latter two correlations were not significant in case of SRT\_Tar. Two multiple linear regression models were built for predicting performance of the BiTDs on SRT\_Id and



**FIGURE 2 |** NWRT: % correct identical repetition (NWRT\_global, NWRT\_LI, and NWRT\_LD).



**FIGURE 3 |** SRT: % identical repetition (SRT\_Id) and correct production of target structure (SRT\_Tar).

**TABLE 5 |** The effect of (clinical) group membership on LITMUS-NWRT and SRT (Kruskal-Wallis tests) and pairwise comparisons (Mann-Whitney U tests).

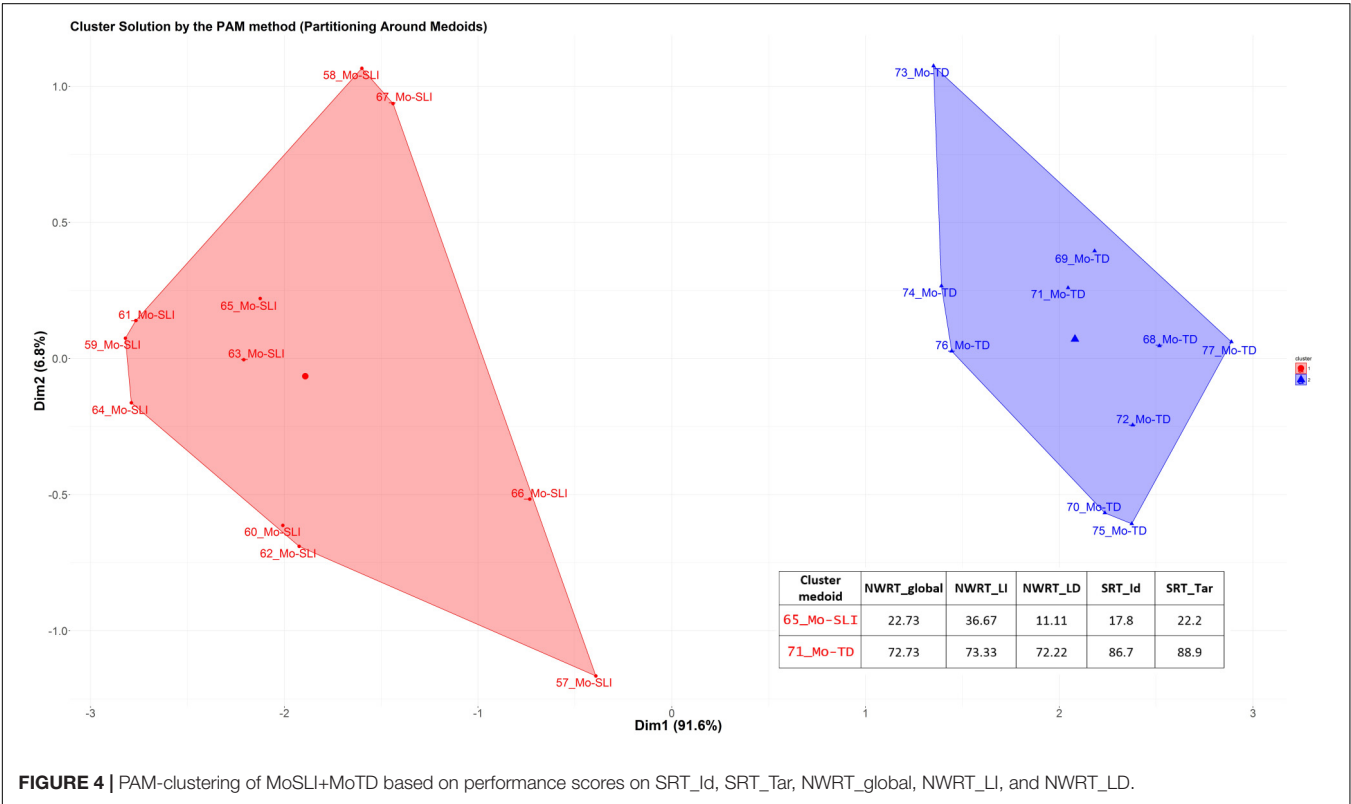
	NWRT_global	NWRT_LI	NWRT_LD	SRT_Id	SRT_Tar
	$\chi^2(5, N = 77)$ = 39.600, $p < 0.001$	$\chi^2(5, N = 77)$ = 35.044, $p < 0.001$	$\chi^2(5, N = 77)$ = 38.714, $p < 0.001$	$\chi^2(5, N = 77)$ = 43.365, $p < 0.001$	$\chi^2(5, N = 77)$ = 43.086, $p < 0.001$
MoTD vs. MoSLI	$U = 0.500$ , $p < 0.001$ , $r = 0.838$	$U = 1.500$ , $p < 0.001$ , $r = 0.824$	$U = 0.000$ , $p < 0.001$ , $r = 0.847$	$U = 0.000$ , $p < 0.001$ , $r = 0.847$	$U = 0.500$ , $p < 0.001$ , $r = 0.846$
BiTD_L1 dom. vs. MoSLI	$U = 2.000$ , $p < 0.001$ , $r = 0.804$	$U = 6.500$ , $p < 0.001$ , $r = 0.775$	$U = 3.000$ , $p < 0.001$ , $r = 0.797$	$U = 33.500$ , $p < 0.001$ , $r = 0.592$	$U = 24.500$ , $p < 0.001$ , $r = 0.652$
BiTD_Balanced vs. MoSLI	$U = 7.000$ , $p < 0.001$ , $r = 0.775$	$U = 8.5000$ , $p < 0.001$ , $r = 0.763$	$U = 10.000$ , $p < 0.001$ , $r = 0.736$	$U = 11.500$ , $p < 0.001$ , $r = 0.717$	$U = 10.000$ , $p < 0.001$ , $r = 0.735$
BiTD_L2 dom. vs. MoSLI	$U = 3.000$ , $p < 0.001$ , $r = 0.845$	$U = 4.500$ , $p < 0.001$ , $r = 0.830$	$U = 3.500$ , $p < 0.001$ , $r = 0.841$	$U = 5.000$ , $p < 0.001$ , $r = 0.824$	$U = 5.000$ , $p < 0.001$ , $r = 0.822$
BiTD_L1_dom. vs. BiSLI	$U = 6.000$ , $p < 0.001$ , $r = 0.765$	$U = 25.000$ , $p < 0.001$ , $r = 0.643$	$U = 8.000$ , $p < 0.001$ , $r = 0.752$	$U = 44.000$ , $p < 0.05$ , $r = 0.512$	$U = 29.500$ , $p < 0.001$ , $r = 0.608$
BiTD_balanced. vs. BiSLI	$U = 2.000$ , $p < 0.001$ , $r = 0.802$	$U = 10.500$ , $p < 0.001$ , $r = 0.688$	$U = 2.000$ , $p < 0.001$ , $r = 0.804$	$U = 3.000$ , $p < 0.001$ , $r = 0.798$	$U = 0.000$ , $p < 0.001$ , $r = 0.829$
BiTD_L2_dom. vs. BiSLI	$U = 9.500$ , $p < 0.05$ , $r = 0.648$	$U = 12.000$ , $p < 0.05$ , $r = 0.616$	$U = 7.000$ , $p < 0.001$ , $r = 0.686$	$U = 4.000$ , $p < 0.001$ , $r = 0.726$	$U = 4.500$ , $p < 0.001$ , $r = 0.718$
BiTD_L1_dom. vs. MoTD	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	$U = 16.500$ , $p < 0.001$ , $r = 0.672$	$U = 21.000$ , $p < 0.001$ , $r = 0.638$
BiTD_L1_dom. vs. BiTD_L2_dom.	<i>n.s.</i>	<i>n.s.</i>	<i>n.s.</i>	$U = 44.500$ , $p < 0.05$ , $r = 0.559$	$U = 53.000$ , $p < 0.05$ , $r = 0.501$

SRT\_Tar. The following variables were entered into the model as independent variables: AoO, LoE, LDI, and SES. The results show that performance on SRT\_Id in the BiTD group is predicted by LDI ( $\beta = 3.724, T = 2.922, p = 0.001$ ), followed by LoE ( $\beta = 3.846, T = 2.287, p = 0.01$ ), and SES ( $\beta = 3.424, T = 2.829, p = 0.001$ ). However, for SRT\_Tar only LDI and SES had significant effects in the full model: LDI ( $\beta = 4.480, T = 3.360, p = 0.001$ ), SES ( $\beta = 2.914, T = 2.301, p = 0.01$ ). The independent variables did not show multi-collinearity in the models ( $VIF < 3$  for all independent variables).

Comparison of global performance of L1-dominant BiTDs to their monolingual, balanced, and L2-dominant peers as well as results of regression analyses show that language dominance was the first predictor to explain performance of the BiTDs on both SRT\_Id and SRT\_Tar, and point to the possibility that language dominance could compromise the diagnostic accuracy of the SRT if administered to bilinguals in their non-dominant language, here German. In order to examine this, we ran the k-medoid PAM-clustering to group the children into SLI vs. TD based on their performance scores on SRT and NWRT, determined the cut-off points between the clusters for each of the repetition measures and calculated the diagnostic accuracy for different combinations of sub-measures of the two. Next, regression analyses using Firth's correction were carried out to examine whether language dominance contributed to results of PAM-clustering, i.e., assigning the children to the clinical vs. non-clinical cluster based on performance scores on the LITMUS repetition tasks. We examined this for SRT and NWRT separately as well as combined. LDI and Positive\_Early\_Development

were entered as predictors for PAM cluster membership into all regression models in addition to two further background variables (age, AoO, LoE, SES).

Before applying the PAM clustering to our bilinguals, we first tested it on our monolingual data set. The following variables were entered in the cluster analysis simultaneously: NWRT\_global, NWRT\_LI, NWRT\_LD, SRT\_Id and SRT\_Tar. The Hopkins statistic yielded a value of around 0.23 indicating clusterable non-random data, and the Gap Statistic revealed that the optimal cluster solution is 2. The clustering procedure resulted in a clear separation into two homogenous groups with two cluster medoids. The cut-off points (see section “Data Analysis”) separating the monolingual clinical cluster from the non-clinical one in our data sample were as follows: SRT\_Id: 40%, SRT\_Tar: 53.3% and NWRT\_global: 45.45%, NWRT\_LI: 60%, NWRT\_LD: 47.22%. **Figure 4** gives a visual representation of the k-medoid PAM-cluster analysis on monolingual data using the two-cluster solution. Cases belonging to the cluster on the right are identified as TD cases, while those in the cluster on the left as SLI cases. To facilitate computing sensitivity and specificity of the task, case numbers were combined with the clinical status as assigned by our classification procedure based on standardized assessment. As can be seen in **Figure 4**, all of the monolingual children assigned to the MoSLI group based on standardized test procedures belong to the clinical cluster, yielding a sensitivity of 100%, whereas all of the monolingual subjects classified as MoTD based on standardized test procedures belonged to the non-clinical cluster, which yields a specificity of 100%. We also ran the clustering procedure



on SRT and NWRT separately, i.e., (SRT\_Id+SRT\_Tar) and (NWRT\_global+NWRT\_LI+NWRT\_LD), respectively and obtained similar results. In a next step, we used regression analysis entering age<sup>23</sup> as a single variable to check whether chronological age could explain cluster membership. Results of the latter analysis indicate that there is no association between age and the cluster variable (Firth:  $\beta = -0.03167$ ,  $Z = -0.646$ ,  $p = 0.519$ ).

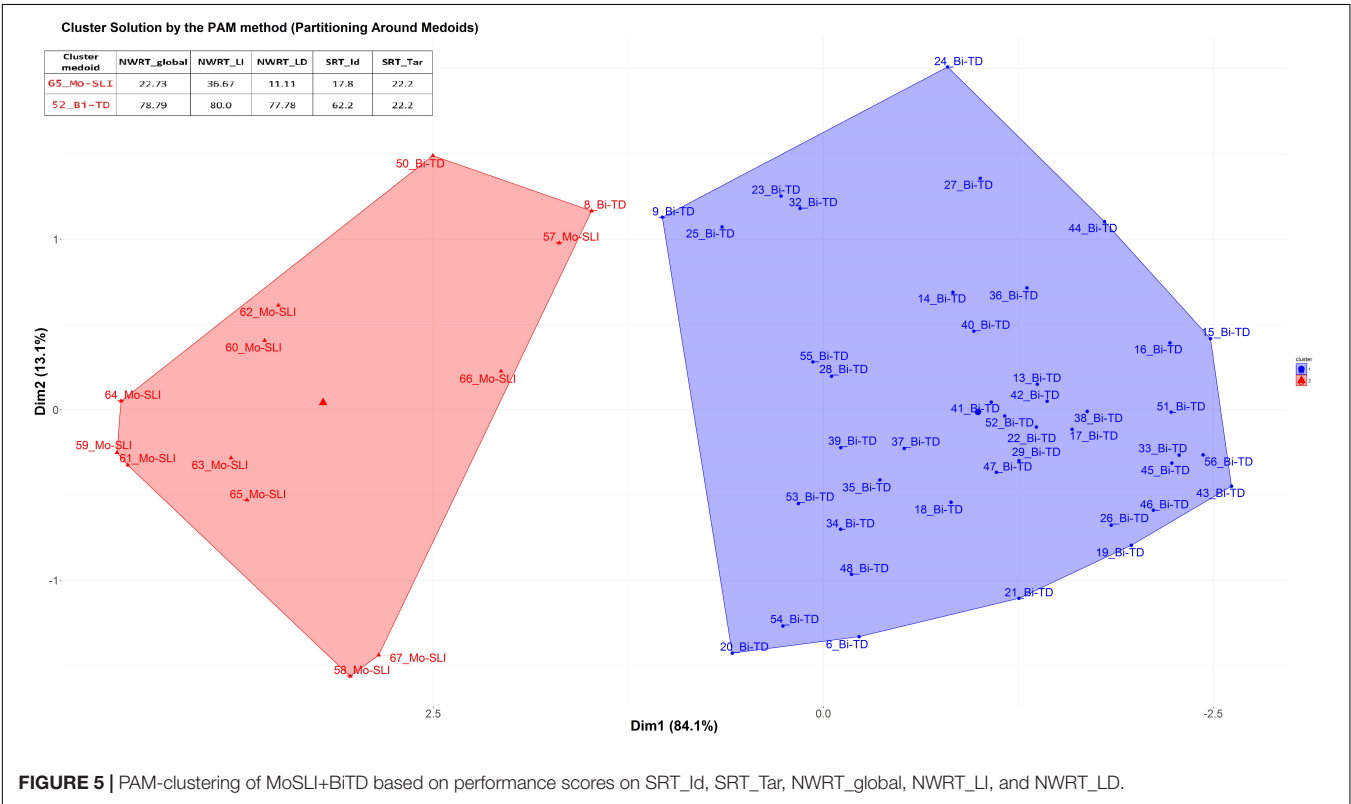
In order to check for overlap between BiTD and MoSLI, we applied the PAM-analysis to the MoSLI and BiTD children collapsed together using performance scores on both SRT and NWRT. The data yielded an H-value of around 0.18, which indicates clusterable non-random data with 2 as the optimal number of clusters. Before entering all five variables into the clustering procedure, we first carried out the clustering procedure based on performance on SRT\_Id+SRT\_Tar. Ctree models showed that SRT\_Tar but not SRT\_Id predicted cluster membership with a threshold of 53.3% separating the two clusters. All MoSLI children scored below cut-off and were thus assigned to the clinical cluster by the PAM algorithm, i.e., sensitivity = 100% with an LR+ = 6.29, whereas 37/44 BiTD children performed above threshold (specificity = 84.1%, LR- = 0.00). Age as a single variable in the regression model did not prove to be a predictor for cluster membership (Firth:  $\beta = -0.02849$ ,  $Z = -1.238$ ,  $p = 0.216$ ). In the next step, we ran the PAM-analysis on NWRT\_global, NWRT\_LI and NWRT\_LD. The clustering resulted in two clusters separated

by a cut-off of 33.33% on NWRT\_LD, which also is the primary predictor of clustering membership. 10/11 MoSLI children were assigned to the clinical cluster, yielding a sensitivity rate of 91% and 43/44 BiTD children performed above cut-off and were assigned to the non-clinical cluster giving a specificity of 98% with an LR+ = 39.56 and LR- = 0.092. Again, age was not a significant predictor for the cluster variable (Firth:  $\beta = -0.03618$ ,  $Z = -1.293$ ,  $p = 0.196$ ).

Finally, both LITMUS-tasks were included in the PAM-analysis using the measures NWRT\_global, NWRT\_LI, NWRT\_LD, SRT\_Id and SRT\_Tar. After entering all SRT and NWRT measures at once into the clustering procedure, 5 of the 7 BiTDs, who were assigned to the clinical cluster based on scores on SRT alone, changed membership from the clinical to the non-clinical cluster. A combination of both SRT and NWRT measures yielded 100% sensitivity (all MoSLIs belong to the clinical cluster) and 95% specificity (42/44 BiTDs belong to the non-clinical cluster) with an LR+ of 20 and an LR- of 0.00. An illustration of the result of the PAM cluster analysis on the MoSLI and BiTD data is given in **Figure 5**. Age at testing as a single variable did not play a significant role in predicting PAM cluster membership (Firth:  $\beta = -0.05145$ ,  $Z = -1.824$ ,  $p = 0.0681$ ).

A visual representation, a ctree, of the hierarchical structure of the most relevant linguistic variables for predicting PAM cluster membership illustrated in **Figure 5** is provided in **Figure 6**. Within a ctree, only those variables serving as relevant to explaining the clustering results appear in the

<sup>23</sup>SES information is only available for the bilingual participants.



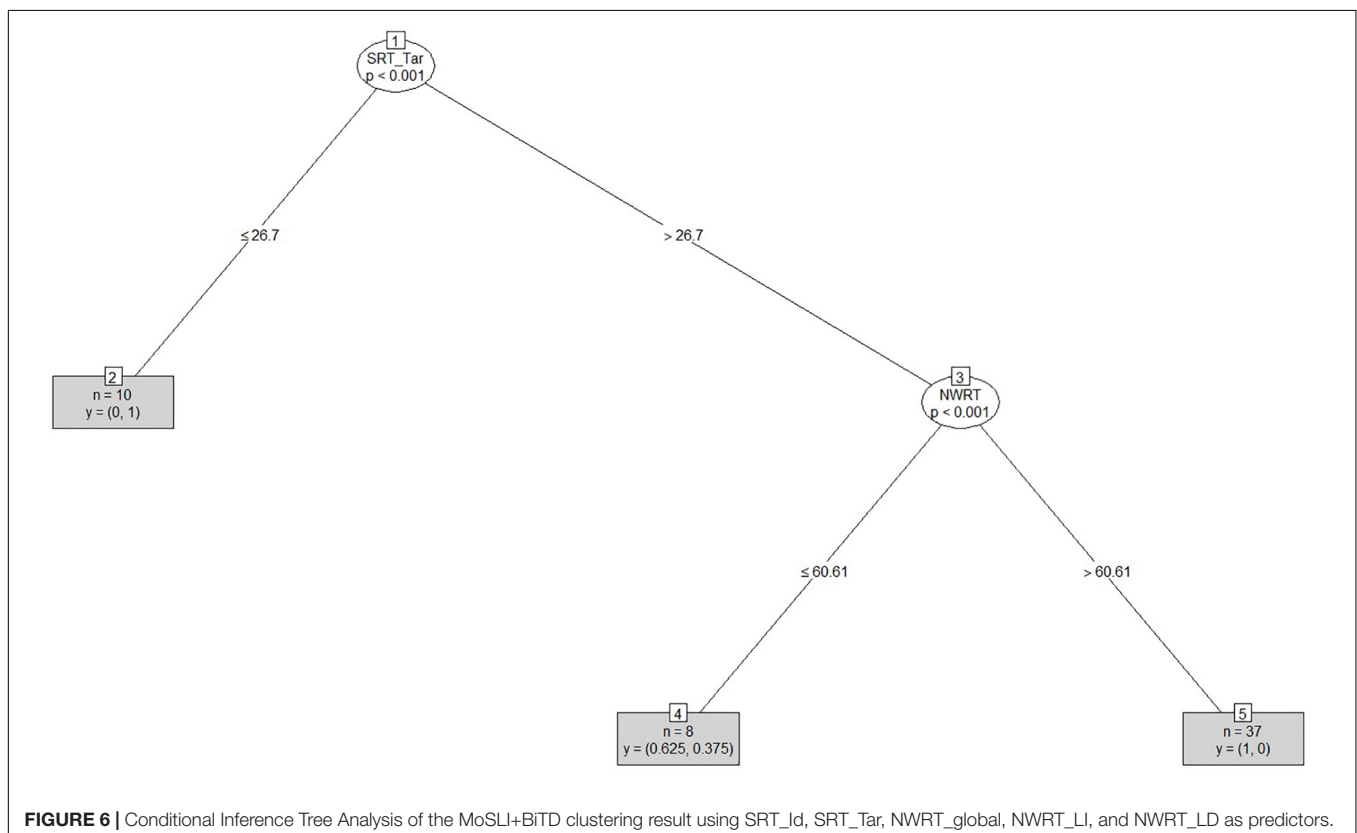


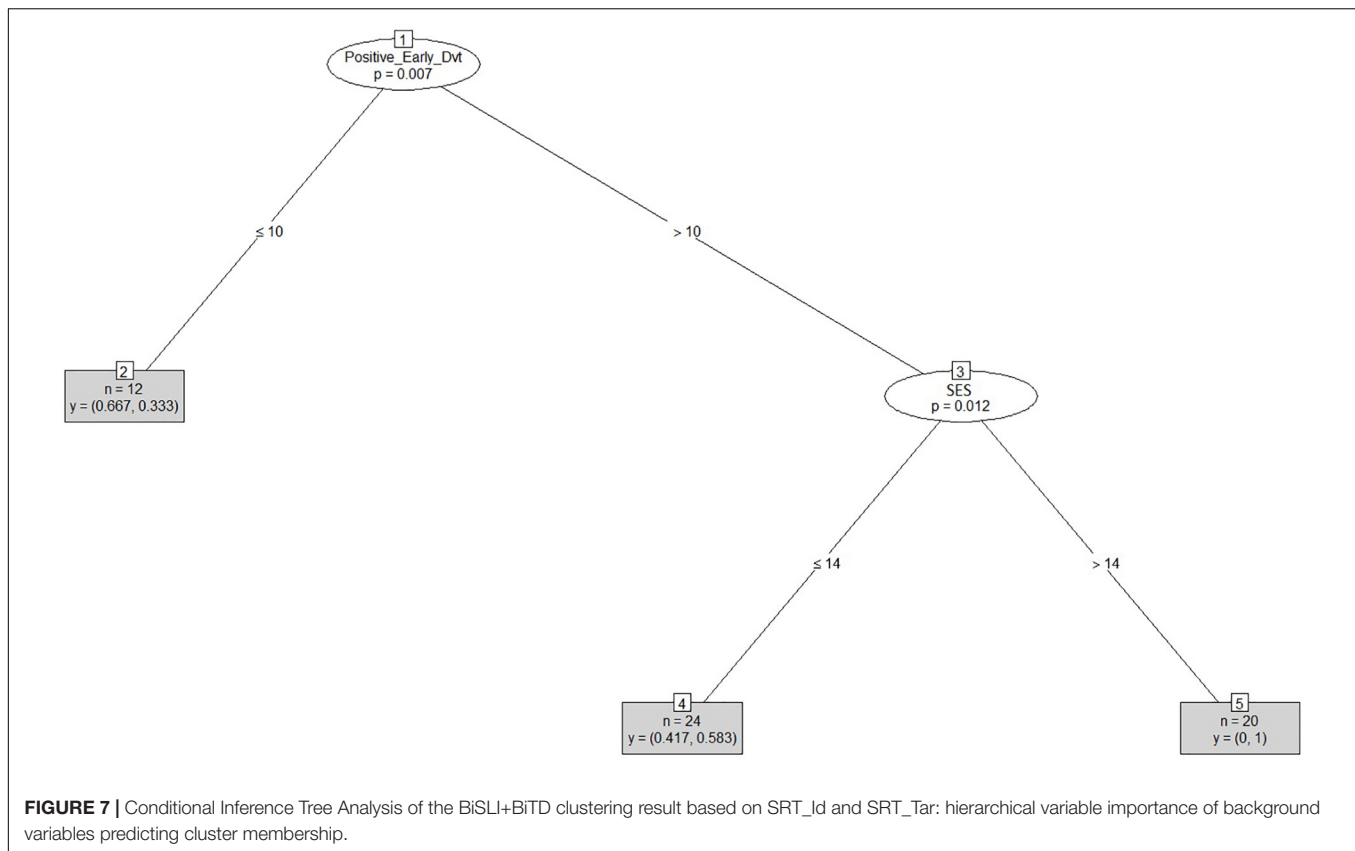
graph, where each relevant variable is represented by an oval circle and classification rules are represented by thresholds. Classification of cases starts at the top node (root). The second most important variable is one level below the top node. Classification then proceeds by moving down the branch until we arrive at a terminal node representing classification accuracy according to PAM clustering<sup>24</sup>, where classification accuracy is represented in squares (y). The two numbers next to “y” show the proportion of cases successfully classified and misclassified as SLI. The number of cases on that route is represented by “n.” Each classification route can be expressed in the form of if-then conditions with cut-offs. As can be seen in **Figure 6**, when all five measures are included in the clustering procedure, both SRT\_Tar and NWRT\_global are identified as significant contributors toward predicting PAM cluster membership. Classification of cases start at the top node occupied by SRT\_Tar followed by the second most important variable “NWRT\_global,” which is one level below the top node. Based on the hierarchical variable structure depicted in the ctree below, it becomes visible that 10 children whose scores on SRT\_Tar were  $\leq 26\%$  were assigned to the clinical cluster. For subjects performing above 26% correct on SRT\_Tar, performance on NWRT\_global was taken into

account giving rise to two roots: (a) if subject performs  $>26\%$  on SRT\_Tar and  $>60.61\%$  on NWRT\_global then assign to non-clinical cluster (TD), (b) if subject performs  $>26\%$  on SRT\_Tar but  $\leq 60.61\%$  on NWRT\_global then assign to clinical cluster (SLI).

Turning now to results of bilingual children, we performed the PAM-analysis on all BiSLI and BiTD groups collapsed together based on the performance scores in the SRT and NWRT. The *Hopkins statistic* indicated regularly spaced data that are neither clustered nor random (H-value of around 0.2) and the Gap statistic suggested the two-cluster solution. Results of the PAM clustering based on performance of BiTDs and BiSLIs on SRT\_Id and SRT\_Tar were similar to those we obtained for BiTDs and MoSLIs (see **Figure 5**). 11/12 BiSLIs were assigned to the clinical cluster yielding a sensitivity of 91.7%, whereas 37/44 BiTDs were assigned to the non-clinical cluster giving a specificity of 84.1%,  $LR+ = 5.76$ ,  $LR- = 0.10$ . The thresholds separating the bilingual clinical cluster from the non-clinical one were 33.3% for SRT\_Id and 53.3% for SRT\_Tar, whereby SRT\_Tar was the main predictor for the clustering result (with the same cut-off of 53.3%). Regression analysis as well as ctree analysis showed that Positive\_Early\_Development (Firth:  $\beta = 1.0636$ ,  $Z = 2.614$ ,  $p = 0.001$ ) followed by SES ( $\beta = 0.7843$ ,  $Z = 2.033$ ,  $p = 0.01$ ) were significant predictors for cluster membership. Variables related to bilingualism, i.e., AoO, LoE and LDI, did not explain cluster membership. An illustration of hierarchical structure of variable importance with classification thresholds is depicted in **Figure 7**.

<sup>24</sup>The classification accuracy given in squares within the ctree refers only to that of PAM clustering and does not represent diagnostic accuracy of cluster membership as measured by comparing cases identified as SLI or TD by the PAM to the clinical status assigned based on our classification procedure outlined in section “Participants.”



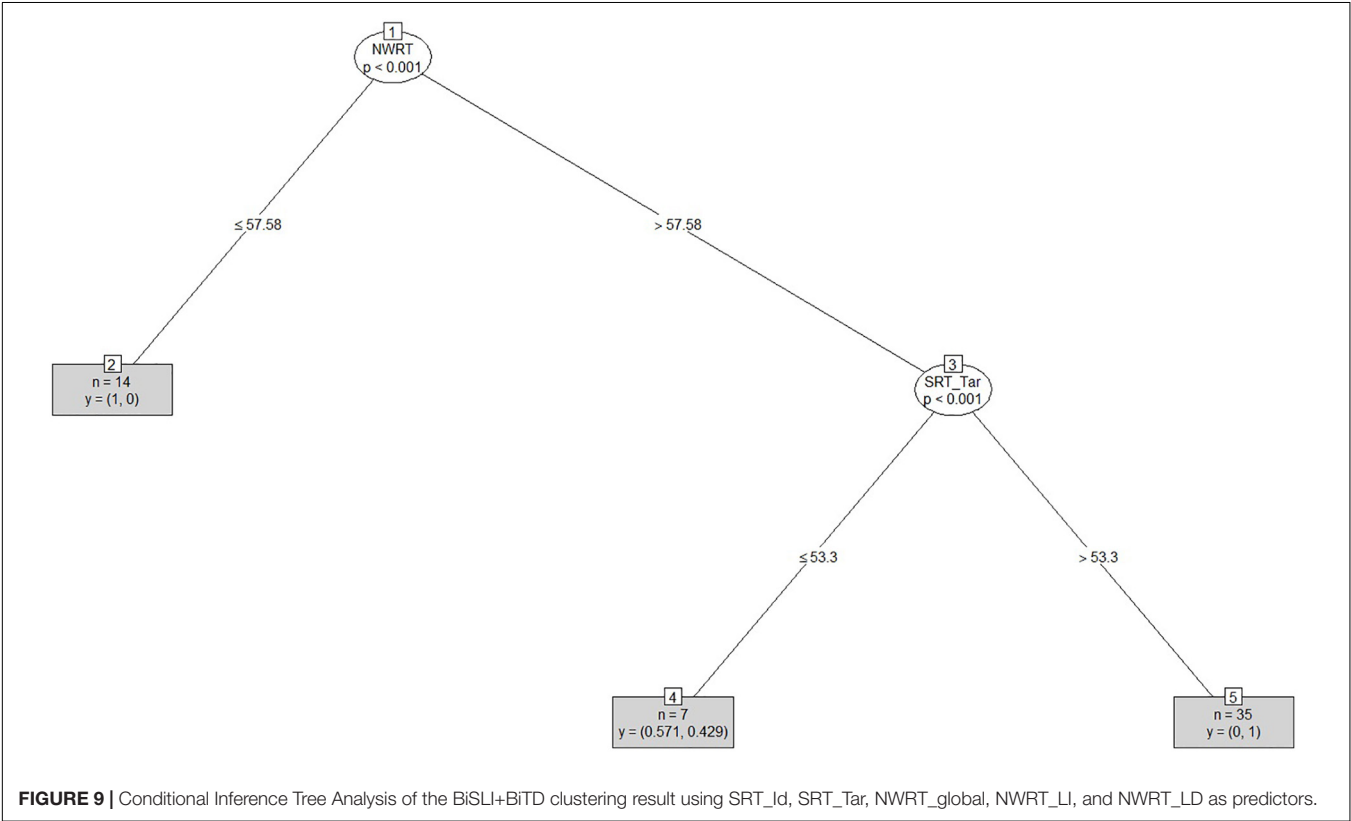
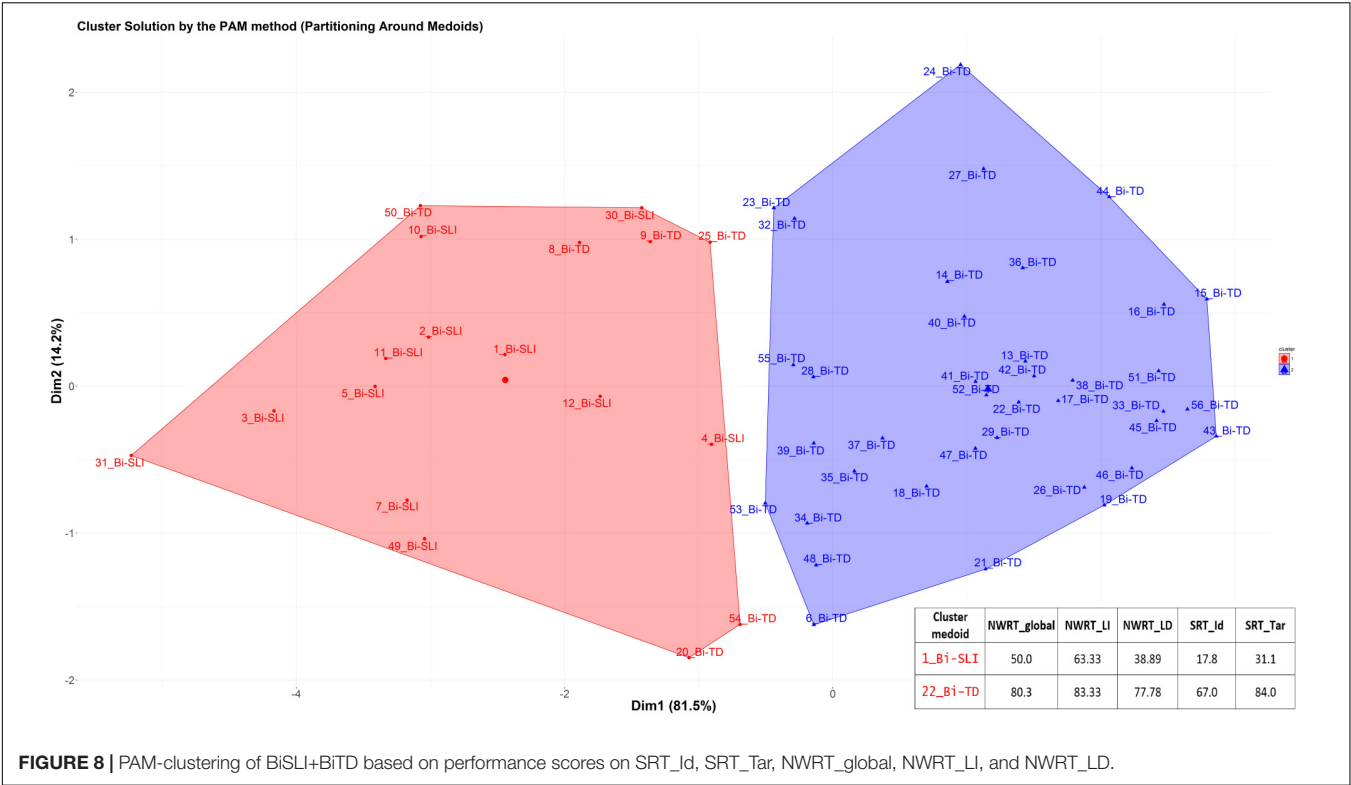


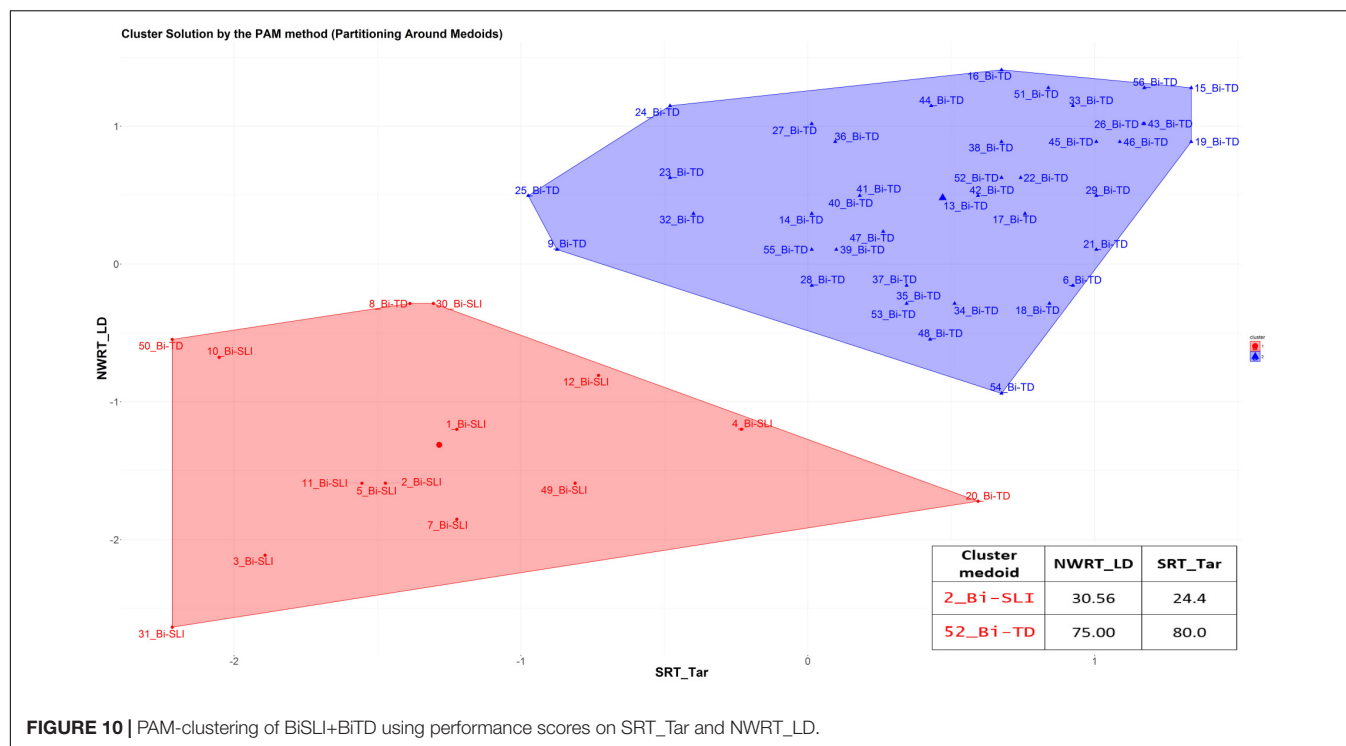
We ran the same clustering procedure on the bilinguals' performance in NWRT using the variables NWRT\_global, NWRT\_LI and NWRT\_LD. All BiSLI children were assigned to the clinical cluster yielding a 100% sensitivity; however, 9 BiTD children were assigned to the clinical cluster, i.e., only 35/44 BiTDs were assigned to the non-clinical cluster (specificity = 80%), LR+ = 5, LR− = 0.00. Ctree analysis showed that NWRT\_global was the main variable predicting cluster membership with a cut-off 66.7%. Next, we ran Firth's biased regression analysis on the clustering results for NWRT entering age, Positive\_Early\_Development, SES and LDI as fixed factors. Results showed that neither language dominance nor SES explained cluster membership based on NWRT\_global. As expected, Positive\_Early\_Development was the main variable explaining the clustering result (Firth:  $\beta = 0.38996$ ,  $Z = 2.626$ ,  $p = 0.001$ ). The other significant predictor for NWRT\_global was chronological age (Firth:  $\beta = 0.05931$ ,  $Z = 2.150$ ,  $p = 0.01$ ).

Since NWRT\_global is a composite score obtained by adding up scores of both of the language independent (NWRT\_LI) and language dependent parts (NWRT\_LD), we wanted to verify whether both of them were affected by the age factor. To achieve this, we ran the PAM-analysis on each of them separately. The results show that if clustering is solely based on performance on NWRT\_LI upon a threshold of 73.3% (as established by ctree analysis), 10 BiTD children would be over-identified as having SLI, yielding a specificity of only 77%,

LR+ = 4.385, LR− = 0.00. Both Positive\_Early\_Development (Firth:  $\beta = 0.38996$ ,  $Z = 2.626$ ,  $p = 0.01$ ) and age (Firth:  $\beta = 0.05591$ ,  $Z = 2.266$ ,  $p = 0.01$ ) were significant predictors for the clustering results (variables entered in the regression model: Positive\_Early\_Development, SES, LDI and age). Ctree analysis showed that the age threshold separating the two clusters based on NWRT\_LI was 87 months (7;3 years). On the other hand, if the bilingual children in our data set are clustered based on performance in NWRT\_LD alone, the diagnostic accuracy drastically improves: upon a 50% cut-off score, only 2/44 BiTD children are assigned to the clinical cluster, while all BiSLI children are classified as SLI, which yields 95% specificity and 100% sensitivity (LR+ = 20, LR− = 0.00). Positive\_Early\_Development was singled out as a predictor explaining cluster membership based on NWRT\_LD (Firth:  $\beta = 0.30611$ ,  $Z = 2.946$ ,  $p = 0.001$ ), i.e., the variables age, LDI and SES did not explain cluster membership. In the following step, we included all SRT and NWRT measures (SRT\_Id, SRT\_Tar, NWRT\_global, NWRT\_LI and NWRT\_LD) in the clustering procedure. As can be seen in **Figure 8**, combining SRT with NWRT enhances the diagnostic accuracy: all of the BiSLI children (12/12) were assigned to the clinical cluster (100% sensitivity), while 39/45 BiTD children were assigned to the non-clinical cluster (87% specificity, LR+ = 7.692, LR− = 0.00).

**Figure 9** shows that both NWRT\_global and SRT\_Tar were significant contributors toward predicting PAM





cluster membership when all 5 variables are included in the clustering procedure. Classification of cases started at top node “NWRT\_global” followed by the second relevant variable “SRT\_Tar,” which is one level below the top node. According to the hierarchical variable structure illustrated in **Figure 9**, 14 children whose scores on NWRT\_global were  $\leq 57.58\%$  were classified as SLI. In case of children with scores above 57.58% on NWRT\_global, performance on SRT\_Tar was taken into consideration leading to two roots: (a) if subject performs  $> 57.58\%$  on NWRT\_global and  $> 53.3\%$  SRT\_Tar, then assign subject to non-clinical cluster (TD), (b) if subject performs  $> 57.58\%$  on NWRT\_global but  $\leq 53.3\%$  on SRT\_Tar then assign to clinical cluster (SLI). Regression analysis using the previous four variables revealed that only Positive\_Early\_Development was a significant predictor for the clustering outcome (Firth:  $\beta = 0.39394$ ,  $Z = 2.907$ ,  $p = 0.001$ ).

To address research question (iii), we ran PAM clustering on bilingual data using scores of NWRT\_LD and SRT\_Tar in order to examine whether a combination thereof yielded the best diagnostic accuracy rates. Indeed, only 2/44 BiTD children were over-identified as having SLI (95% specificity) and all of the 12 BiSLI children were assigned to the clinical cluster (100% sensitivity) with an LR+ of 20 and an LR- of 0). The cut-off scores were 52.78% for NWRT\_LD and 53.3% for SRT\_Tar, with NWRT LD being the primary predictor for clustering results followed by SRT\_Tar. Only Positive\_Early\_Development was a significant predictor of cluster membership (Firth:  $\beta = 0.39394$ ,  $Z = 2.907$ ,  $p = 0.001$ ). The clustering results are depicted in **Figure 10**.

## DISCUSSION AND CONCLUSION

The purpose of this study was to evaluate the robustness of two LITMUS tools, German LITMUS-SRT and NWRT, against the influence of language dominance on their diagnostic accuracy for SLI in bilingual children. Since both tasks were designed to minimize bias against bilingual populations while being indicative of the presence of LI, we wanted to specifically verify whether the tasks were only sensitive to risk factors for SLI or whether background variables related to bilingualism, particularly, the degree of language dominance (as measured by relative amount of use and exposure to L1/L2) could influence the performance of BiTDs to an extent that would compromise their diagnostic accuracy. The second aim of the study was to investigate whether combining LITMUS-SRT (especially when scored by correct target structure) with NWRT yielded better diagnostic accuracy than single measures and helped avoid cases of misdiagnosis. Following our own research (e.g., Abed Ibrahim and Hamann, 2017; Hamann and Abed Ibrahim, 2017; Grimm and Hübner, in press), we particularly wanted to check whether a combination of German SRT\_Tar and the language dependent part of the NWRT yielded higher diagnostic accuracy for identifying SLI in bilingual children than other combinations of measures. The former was found to be a fairer method than identical repetition for scoring SRT as it compensates for typical L2-errors such as lexical substitutions, while the latter was shown to maximize the performance gap between SLI and TD not only in monolinguals but also in bilinguals given its higher level of structural complexity.

In order to examine this, we first compared global performance of L1-dominant BiTDs to that of MoTDs, balanced



and L2-dominant-BiTDs as well as to MoSLIs and BiSLIs. Results showed that although all three BiTD groups (regardless of their dominance) significantly outperformed MoSLIs and BiSLIs on all SRT and NWRT measures, L1-dominant-BiTDs were significantly outperformed by MoTDs and L2-dominant-BiTDs on both SRT\_Id and SRT\_Tar with large effect sizes (see **Figures 2, 3**). This echoes the findings of Meir (2018), who reported similar results for performance of Russian-Hebrew bilinguals on LITMUS-SRTs in their weaker heritage or societal language. Our results further showed that the performance gap between monolingual and bilingual SLI and TD groups was larger for NWRT\_LD as opposed to the structurally less complex language independent part of the NWRT and the composite score of the two parts “NWRT\_global” (see **Figure 2**). This is in line with previous work showing that the complexity factors involved in the NWRT\_LD part (i.e., presence of trilateral onset clusters and /sC/ clusters violating the Sonority Sequencing Principle) is particularly challenging for language impaired children regardless of lingual status (Ferré et al., 2015; dos Santos and Ferré, 2018; Grimm and Hübner, in press).

Since language dominance was used as a categorical variable to classify BiTDs in our between-group comparisons, we had to entertain the possibility that the assumed dominance effect for L1-dominant children might have been caused by confounding variables such as age of onset of exposure to L2 (AoO), length of exposure to L2 (LoE) and SES. As for SRT\_Id and SRT\_Tar, we found moderate correlations between performance and language dominance as well as SES, in addition to weak correlations for SRT\_Id with LoE and AoO. Regression analysis showed that language dominance was the key predictor explaining variance in the performance of the BiTDs on SRT\_Id and SRT\_Tar followed by SES. That AoO and LoE did not predict performance of BiTDs on the SRT was an expected outcome since the vast majority of the participants in our bilingual sample were either simultaneous or early successive and were exposed to German for at least 24 months at the time of assessment (see also Armon-Lotem, 2011 for similar results on L2-Hebrew-SRT).

The finding that language dominance influenced the performance of BiTDs on both measures of the LITMUS-SRT questioned its applicability for the identification of SLI in L1-dominant children when administered in their weaker language (German). To answer this, we used a prominent unsupervised machine learning technique, the Partitioning Around Medoids (PAM) for establishing an automatic classification of the monolingual and bilingual children in our data set as TD vs. SLI directly from their performance scores on SRT and NWRT without using information about their clinical status. Subsequently, we compared the participants' clinical group membership revealed by PAM-clustering to their clinical status based on standardized assessment in L1/L2, and calculated sensitivity and specificity (diagnostic accuracy) levels of the tasks in isolation and combined. We also explored which combinations of the measures obtained from SRT\_Id, SRT\_Tar, NWRT\_global, NWRT\_LI, and NWRT\_LD yielded the highest diagnostic accuracy. Finally, we conducted regression analysis to investigate whether background variables other than risk factors for SLI, in particular language dominance (LDI),

explained PAM-cluster membership as TD or SLI based on performance scores on SRT and/or NWRT. Since the index of Positive\_Early\_Development was shown to be a strong predictor for SLI in bilinguals (Boerma and Blom, 2017; Tuller et al., 2018), our premise was that if PAM-cluster membership can only be predicted by this index and not by language dominance or other background variables known to influence performance on repetition tasks (age, AoO, LoE, SES), then clustering of cases cuts across the SLI/TD dimension confirming that the LITMUS-SRT and NWRT are only sensitive to the presence of SLI and are not biased against bilingual children, who are non-dominant in the societal language.

In Hamann and Abed Ibrahim (2017), unsupervised (clustering) machine learning algorithms were only applied to the bilingual data, while Receiver Operating Characteristic curve (ROC) analysis was used to calculate sensitivity and specificity levels for the monolingual data. Given that ROC analysis uses “clinical status” (as assigned by standardized test procedures) as a dependent variable for predicting the sensitivity and specificity of a test, we wanted to verify this finding for the monolinguals using a method independent of “clinical status.” PAM-clustering solely based on scores in SRT\_Id, SRT\_Tar, NWRT\_global, NWRT\_LI, and NWRT\_LD yielded even higher diagnostic accuracy than that in Hamann and Abed Ibrahim (2017). The fact that all of the subjects identified as MoSLI by standardized assessment belonged to the lower performing cluster, while all of the MoTDs belonged to the higher performing cluster (100% sensitivity and 100% specificity) provides additional evidence that these linguistically motivated tasks are very sensitive to the presence of LI in monolinguals and tap the core morphosyntactic and phonological deficits in SLI. The source of the improved diagnostic accuracy as compared to results based on ROC-analysis in Hamann and Abed Ibrahim (2017) is most likely the simultaneous inclusion of both tasks into the clustering procedure and the lower cut-off points obtained by applying ctree to the PAM clustering. This is reminiscent of Armon-Lotem and Meir's (2016) study, which reported an increase in diagnostic accuracy when LITMUS-SRT is supplemented by NWRT for Hebrew and Russian monolinguals. A further important result was that chronological age could not predict cluster membership for the age range in our monolingual data set.

After establishing that both LITMUS-SRT and NWRT were sensitive to SLI in monolinguals, we proceeded to address the frequently reported overlap between MoSLI and BiTD children (e.g., Håkansson and Nettelbladt, 1996; Armon-Lotem, 2010; Paradis, 2010; Hamann, 2012). PAM-clustering conducted on SRT scores entering both measures SRT\_Id and SRT\_Tar yielded good overall diagnostic accuracy (100% sensitivity and 84.1%<sup>25</sup> specificity) with SRT\_Tar being the leading variable for predicting cluster membership since it led to a better separation between

<sup>25</sup>It is worthwhile mentioning that three of the 7 BiTD children assigned to the clinical cluster started acquiring L2 German after the age of five and spent their first years in an exclusive L1-environment. The latter children demonstrated high rates of determiner errors, especially case and gender errors which could affect the realization of target structure of a sizable proportion of the test items of the LITMUS-SRT (see also Abed Ibrahim et al., 2018).

the BiTD and MoSLI clusters. Several studies found this scoring method better suited for assessing morphosyntactic abilities in bilingual children, since it only focuses on the mastery of syntactic structure and does not penalize bilingual children for frequent L2-errors such as lexical substitutions (Armon-Lotem and Meir, 2016; Hamann and Abed Ibrahim, 2017; Hamann et al., 2017; Abed Ibrahim et al., 2018; Meir, 2018).

Next, we checked whether the overlap problem between MoSLI and BiTD could be overcome by using SRT in combination with NWRT. Indeed, including NWRT scores into the clustering procedure resulted in much better diagnostic accuracy with almost no overlap between MoSLI and BiTD (100% sensitivity and 95% specificity). As also reported in Armon-Lotem and Meir (2016), de Almeida et al. (2017), Boerma and Blom (2017), and Hamann and Abed Ibrahim (2017), the latter finding corroborates that a combination of LITMUS instruments assessing different areas of language ability helps to avoid cases of misdiagnosis. Among the five measures SRT\_Id, SRT\_Tar, NWRT\_global, NWRT\_LI and NWRT\_LD, both SRT\_Tar and NWRT\_global were main predictors for clustering results with SRT\_Tar being the more important contributor (see **Figure 6**). We further demonstrated that chronological age did not predict cluster membership here either.

As to the diagnosis of bilinguals, PAM clustering based on scores in SRT\_Id and SRT\_Tar resulted in good overall accuracy rates (91.7% sensitivity and 84.1% specificity). Interestingly, the same 7 BiTDs previously assigned to the clinical cluster upon comparison with MoSLIs were classified as SLI by the PAM as well showing that changing the reference group had no influence on the individual classification of the BiTDs. Again, SRT\_Tar, which compensates for L2-errors, was the primary contributor toward the clustering results with a cut-off 53.3%, which is very close to the threshold obtained by *k*-means clustering in Hamann and Abed Ibrahim (2017). Of the five background variables considered for regression analysis, just two variables unrelated to bilingualism emerged as significant predictors for clustering membership: Positive\_Early\_Development followed by SES. The influence of language dominance, which was a significant predictor explaining the variance in the performance of the BiTDs in SRT\_Id and SRT\_Tar, was outweighed by the presence of risk factors for SLI and was rendered insignificant once the BiSLIs became part of the equation. This is consistent with the findings of Tuller et al. (2018), who found for the German children that Positive\_Early\_Development was the leading predictor for performance in SRT (followed by SES) over variables related to bilingualism.

The clustering solution based on NWRT\_global, NWRT\_LI and NWRT\_LD scores yielded only fair diagnostic accuracy rates due to reduced specificity (specificity = 80%). NWRT\_global emerged as the main predictor for clustering results. Regression analysis revealed that not only Positive\_Early\_Development (most important predictor) but also chronological age were significant predictors for clustering results based on performance scores in NWRT\_global. Given that NWRT\_global is a composite score computed by adding up performance scores in NWRT\_LI, and NWRT\_LD, and since Grimm and Hübner (in press) reported an overlap between MoSLI and BiTD on NWRT\_LI

and better discriminatory power for NWRT\_LD in children aged 8;0 to 10;0 years, we ran cluster analyses on both subparts of the NWRT separately to check for age effects. The analysis revealed that in addition to Positive\_Early\_Development, cluster-membership based on NWRT\_LI was predicted by chronological age with a threshold of 7;3 years, whereas cluster-membership based on NWRT\_LD was not predicted by age and was only sensitive to risk factors for SLI. On the other hand, neither bilingualism related factors nor SES predicted cluster membership derived by performance scores on NWRT\_global or subtests thereof. The latter result echoes what has been found for this type of NWRT in de Almeida et al. (2017) as well as in Tuller et al. (2018).

We have also shown that including all five SRT and NWRT measures in the clustering procedure enhances diagnostic accuracy for SLI in bilingual children, where NWRT\_global and SRT\_Tar were the main contributors explaining the results of the cluster solution. Interestingly, once SRT is combined with NWRT, only Positive\_Early\_Development emerges as a significant predictor for clustering results and SES does not play a role anymore, which is in line with the findings of Chiat and Polišenská (2016).

Given that clustering by scores on NWRT\_LI appeared to be influenced by age, while NWRT\_LD was only sensitive to risk factors of SLI (Positive\_Early\_Development) and since the SRT\_Tar was the chief contributor toward clustering results when both SRT\_Id and SRT\_Tar were included in any clustering procedure on bilingual performance, we expected a combination of SRT\_Tar and NWRT\_LD to yield better diagnostic accuracy rates than other combinations of measures. Indeed, clustering based on performance scores on SRT\_Tar and NWRT\_LD yielded the highest diagnostic accuracy, where only Positive\_Early\_Development predicted clustering results. The crucial contribution of the structurally more complex NWRT\_LD toward diagnostic accuracy is consistent with the robust effects of phonological complexity found in the respective studies (e.g., Gallon et al., 2007; Ferré et al., 2012), with clinical implications that phonological complexity can be used as a reliable indicator for SLI in both monolingual and bilingual children (see Grimm and Hübner, in press). Our results concerning the NWRT\_LD part might seem at odds with results of other studies showing better diagnostic accuracy for *Crosslinguistic*-NWRTs over *Language-Specific*-NWRTs in bilingual populations, e.g., Boerma et al. (2015), Armon-Lotem and Meir (2016), and Boerma and Blom (2017). This can clearly be ascribed to differences in the construction of the tasks, which, as described in the section “The German LITMUS Nonword Repetition Task”, tap different aspects vulnerable in SLI (i.e., phonological working memory vs. phonological complexity), and differ considerably from each other, especially in their language dependent parts. Another possible reason for the poor diagnostic accuracy reported for the *Language-Specific*-NWRTs in the latter three studies might be relatively young age of their participants (5;0–6;0) compared to the age range in our sample (5;6–9;0), which covers the last year of preschool and the first 2–3 primary school years. A study by Rispens and Baker (2012) demonstrated that

both lexical knowledge and discrimination ability significantly influenced performance on NWRT in 5-year-old MoTDs, while this kind of relation could not be attested for 8-year olds.

In line with our previous research, the results presented here and the fact that they emerge from unsupervised PAM-clustering clearly indicate that the German LITMUS- SRT and NWRT are promising tools for the identification of LI in bilingual populations with diverse dominance profiles. We replicated the finding that SRT\_Tar is better suited than SRT\_Id for the assessment of language abilities of bilingual children with German as L2 on a slightly larger group of children with a statistical method better suited for our data set. Even though dominance influences the performance of BiTDs, especially in the SRT, we demonstrated that the diagnostic accuracy of these tools is not compromised by language dominance: while risk factors for SLI were significant predictors for clinical status in all models, language dominance did not contribute at all to explaining results of any of the clustering procedures. Moreover, our results confirmed that using a combination of tasks, each emphasizing a different aspect of language ability, enhances diagnostic accuracy and helps avoid cases of misdiagnosis. As a last promising result, we showed that using SRT\_Tar in conjunction with NWRT\_LD renders the best diagnostic accuracy so far obtained in studies on similarly constructed tasks, where the combination of measures is only sensitive to risk factors for SLI, but not to language dominance nor to SES, which is not achieved by many tasks. We therefore feel confident in pursuing these investigations in order to be able to provide useful and easy to administer L2-tools for clinical use in bilingual contexts. Finally, it should be noted that vast majority of the bilingual children in our sample were either simultaneous or early successive bilinguals, who had at least 2 years of exposure to the L2. Thus,

future research should focus on testing the applicability of this particular combination of tasks to bilinguals with less exposure to the L2.

## AUTHOR CONTRIBUTIONS

LAI developed the theoretical framework, gathered and evaluated the data. IF conducted the statistical analysis. All authors wrote the manuscript.

## FUNDING

The BiLaD project (Bilingual Language Development: Typically Developing Children and Children with Specific Language Impairment) was financed by a joint grant (German DFG: HA 2335/6-1, CH 1112/2-1, and RO 923/3-1) and French ANR grant (ANR-12-FRAL-0014-01) to Laurice Tuller and her team.

## ACKNOWLEDGMENTS

We deeply thank Cornelia Hamann for her continued support and constructive comments. We are grateful to Solveig Chilla, Monika Rothweiler, Hilal Şan, Tatjana Lein, and our French BiLaD partners for their support with data collection and analysis. Our special thanks go to the participating children and their parents as well as to the collaborating kindergartens and speech-language therapy centers. We also thank the three reviewers for their constructive comments and suggestions on previous drafts of the manuscript.

## 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, MA: Cascadia Press), 1–17.
- Abed Ibrahim, L., Hamann, C., and Öwerdieck, D. (2018). "Identifying specific language impairment (SLI) across different bilingual populations: German sentence repetition task (SRT)," in *Proceedings of the 42nd Annual Boston University Conference on Language Development*, eds A. B. Bertolini and M. J. Kaplan (Somerville, MA: Cascadia Press), 1–14.
- American Speech-Language-Hearing Association [ASHA] (2004). *Knowledge and Skills Needed by Speech-Language Pathologists and Audiologists to Provide Culturally and Linguistically Appropriate Services [Knowledge and Skills]*. Available at: [www.asha.org/policy](http://www.asha.org/policy)
- Archibald, L. M., and Gathercole, S. E. (2006). Nonword repetition: a comparison of tests. *J. Speech Lang. Hear. Res.* 49, 970–983. doi: 10.1044/1092-4388(2006/070)
- Archibald, L. M., and Gathercole, S. E. (2007). The complexities of complex memory span: storage and processing deficits in specific language impairment. *J. Mem. Lang.* 57, 177–194. doi: 10.1016/j.jml.2006.11.004
- Argyri, E., and Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Biling. Lang. Cogn.* 10, 79–99. doi: 10.1017/S1366728906002835
- Armon-Lotem, S. (2010). Instructive bilingualism: can bilingual children with specific language impairment rely on one language in learning a second one? *Appl. Psycholinguist.* 31, 253–260. doi: 10.1017/S0142716409990385
- Armon-Lotem, S. (2011). Design and data of English-Hebrew and Russian-Hebrew sentence repetition tasks. *Paper Presented at COST IS0804 WG and MC Meeting*, Malta.
- Armon-Lotem, S., and de Jong, J. (2015). "Introduction," in *Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters). doi: 10.21832/9781783093137
- Armon-Lotem, S., de Jong, J., and Meir, N. (eds). (2015). *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*. Bristol: Multilingual Matters. doi: 10.21832/9781783093137
- 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
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2
- Balladares, J., Marshall, C., and Griffiths, Y. (2016). Socio-economic status affects sentence repetition, but not non-word repetition, in Chilean preschoolers. *First Lang.* 36, 338–351. doi: 10.1177/0142723715626067
- Banerjee, A., and Davé, R. N. (2004). "Validating clusters using the Hopkins statistic," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, Piscataway, NJ, 149–153. doi: 10.1109/FUZZY.2004.1375706
- Barlow, J. A. (2001). The structure of /s/ sequences: evidence from a disordered system. *J. Child Lang.* 28, 291–324. doi: 10.1017/S0305000901004652
- Bedore, L. M., and Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: current findings and implications for practice. *Int. J. Biling. Educ. Biling.* 11, 1–29. doi: 10.2167/beb392.0



- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Green, K., et al. (2012). The measure matters: language dominance profiles across measures in Spanish-English bilingual children. *Biling. Lang. Cogn.* 15, 616–629. doi: 10.1017/S1366728912000090
- 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
- Bishop, D. V., Adams, C. V., and Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes Brain Behav.* 5, 158–169. doi: 10.1111/j.1601-183X.2005.00148.x
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., and the Catalise-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *J. Child Psychol. Psychiatry* 58, 1068–1080. doi: 10.1111/jcpp.12721
- 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., Leleman, 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
- Bortolini, U., and Leonard, L. (2000). Phonology and children with specific language impairment: the status of structural constraints in two languages. *J. Commun. Disord.* 33, 131–150. doi: 10.1016/S0021-9924(99)00028-3
- Bulheller, S., and Häcker, H. (2002). *Coloured Progressive Matrices*. Frankfurt: Pearson Assessment.
- Calvo, A., and Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition* 130, 278–288. doi: 10.1016/j.cognition.2013.11.015
- Cantone, K. F., Müller, N., Schmitz, K., and Kupisch, T. (2008). Rethinking language dominance in bilingual children. *Linguistische Berichte* 215, 307–343.
- 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.
- Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36. doi: 10.18637/jss.v061.i06
- 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., Armon-Lotem, S., Marinis, T., Polišenská, K., Roy, P., and Seeff-Gabriel, B. (2013). “The potential of sentence imitation tasks for assessment of language abilities in sequential bilingual children,” in *Issues in the Assessment of Bilinguals*, ed. V. Mueller-Gathercole (Bristol: Multilingual Matters), 56–89.
- Chiat, S., and Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: effects of bilingualism and socioeconomic status on children’s performance. *J. Speed. Lang. Hear. Res.* 59, 1179–1189. doi: 10.1044/2016\_JSLHR-L-15-0293
- Chilla, S., Hamann, C., Prévost, P., Abed Ibrahim, L., Ferré, S., dos Santos, C., et al. (in press). “The influence of different first languages on L2 LITMUS-NWR and L2 LITMUS-SRT in French and German: a crosslinguistic approach”, in *LITMUS in Action: Comparative Studies across Europe, TILAR*, eds K. Grohmann and S. Armon-Lotem (Amsterdam: John Benjamins Publishing Company).
- Chilla, S., and Şan, H. (2017). “Möglichkeiten und Grenzen der Diagnostik ersprachlicher Fähigkeiten: Türkisch-deutsche und türkisch-französische Kinder im Vergleich,” in *Sprachen 2016. Russisch und Türkisch im Fokus*, eds C. Yildiz et al. (Berlin: Peter Lang), 175–205.
- Chomsky, N. (1986). *Knowledge of Language: its Nature, Origin and Use*. New York, NY: Praeger.
- Chomsky, N. (2005). Three factors in language design. *Linguist. Inq.* 36, 1–22. doi: 10.1162/0024389052993655
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguist. Approaches Biling.* 1, 318–345. doi: 10.1075/lab.1.3.05cho
- Clahsen, H. (1991). *Child Language and Developmental Dysphasia: Linguistic Studies of the Acquisition of German*. Amsterdam: John Benjamins. doi: 10.1075/sspcl.2
- Coady, J., and Evans, J. (2008). Uses and interpretations of nonword repetition tasks in children with and without specific language impairments. *Int. J. Lang. Commun. Disord.* 43, 1–40. doi: 10.1080/13682820601116485
- 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
- Crutchley, A., Botting, N., and Conti-Ramsden, G. (1997). Bilingualism and specific language impairment in children attending language units. *Int. J. Lang. Commun. Disord.* 32, 267–276. doi: 10.3109/13682829709020409
- Dalal, D., and Zickar, M. (2011). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organ. Res. Methods* 15, 339–362. doi: 10.1177/1094428111430540
- de 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, 331–358. doi: 10.1075/lab.15019.alm
- de Houwer, A. (2009). *Bilingual First Language Acquisition*. Bristol: Multilingual Matters.
- de Houwer, A., and Bornstein, M. H. (2016). “Balance patterns in early bilingual acquisition: a longitudinal study of word comprehension and production,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 134–155. doi: 10.1017/CBO9781107375345.007
- Deuchar, M., and Muntz, R. (2003). “Factors accounting for code-mixing in an early developing bilingual,” in *(In)vulnerable Domains in Multilingualism*, ed. N. Müller (Amsterdam: John Benjamins), 161–190.
- Dollaghan, C., and Horner, E. (2011). Bilingual language assessment: a meta-analysis of diagnostic accuracy. *J. Speech Lang. Hear. Res.* 54, 1077–1088. doi: 10.1044/1092-4388(2010/10-0093)
- Döpke, S. (2000). *One Parent-One Language: An Interactional Approach*. Amsterdam: Benjamins.
- dos Santos, C., and Ferré, S. (2018). A non-word repetition task to assess bilingual children’s phonology. *Lang. Acquis.* 25:1. doi: 10.1080/10489223.2016.1243692
- Eisenberg, P. (1999). *Der Satz. Grundriß der deutschen Grammatik*. Stuttgart: J. B. Metzler. doi: 10.1007/978-3-476-03765-7
- Engel, P. M. J., Santos, F. H., and Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *J. Speech Lang. Hear. Res.* 51, 1580–1587. doi: 10.1044/1092-4388(2008/07-0210)
- Engel de Abreu, P. M. J., Baldassi, M., Puglisi, M. L., and Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: testing language minority children with an immigrant background. *J. Speech Lang. Hear. Res.* 56, 630–642. doi: 10.1044/1092-4388(2012/12-0079)
- Ferré, S., dos Santos, C., and de Almeida, L. (2015). “Potential Phonological Markers for SLI in Bilingual Children,” in *Proceedings of the 39th Annual Boston University Conference on Language Development* (Somerville, MA: Cascadia Press), 152–164.
- 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. Hooley, L. Bombine, M. Pouplier, C. Mooshammer, and B. Kühnert (Berlin: de Gruyter), 285–308. doi: 10.1515/9781614510772.285
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27–38. doi: 10.1093/biomet/80.1.27
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., and Zebib, R. (2018). How to identify SLI in bilingual children: a study on sentence repetition in French. *Lang. Acquis.* 25:1. doi: 10.1080/10489223.2016.1192635
- Flores, C. (2015). Losing a language in childhood: a longitudinal case study on language attrition. *J. Child Lang.* 42, 562–590. doi: 10.1017/S0305000914000233
- Fox-Boyer, A. V. (2014). *Psycholinguistischer Analyse kindlicher Aussprachestörungen-II: PLAKSS II*. Frankfurt: Pearson.



- Fredman, M. (2006). Recommendations for working with bilingual children – Prepared by the multilingual affairs committee of IALP. *Folia Phoniatr. Logop.* 58, 458–464.
- 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
- Friedmann, N., and Novogrodsky, R. (2008). “Subtypes of SLI: SySLI, PhoSLI, LeSLI, and PraSLI”, in *Language Acquisition and Development*, eds A. Gavarro and M. J. Freitas (Cambridge: Cambridge University), 205–217.
- 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 Sztetman, 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. doi: 10.1075/la.223.14fri
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, 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. doi: 10.21832/9781783093137-011
- 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
- Gathercole, S. (2006). Complexities and constraints in nonword repetition and word learning. *Appl. Psycholinguist.* 27, 599–613. doi: 10.1017/S014271640606053X
- Gathercole, S., and Baddeley, A. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *J. Mem. Lang.* 29, 336–360. doi: 10.1016/0749-596X(90)90004-J
- Gathercole, V. C. M. (2016). “Factors moderating proficiency in bilingual speakers,” in *Bilingualism Across the Lifespan: Factors Moderating Language Proficiency*, eds E. Nicoladis and S. Montanari (Washington DC: De Gruyter Mouton and American Psychological Association), 123–140. doi: 10.1037/14939-008
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Biling. Lang. Cogn.* 12, 213–237. doi: 10.1017/S1366728909004015
- Genesee, F., and Nicoladis, E. (2007). “Bilingual first language acquisition,” in *Blackwell Handbook of Language Development*, eds E. Hoff and M. Shatz (Oxford: Blackwell Publishing Ltd.), 324–342. doi: 10.1002/9780470757833.ch16
- Genesee, F., Paradis, J., and Crago, M. (2004). *Dual Language Development and Disorders: a Handbook on Bilingualism and Second Language Learning*. Baltimore: Brookes Publishing.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Glück, C. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-Jährige: WWT 6–10*, 2nd Edn. Munich: Elsevier, Urban & Fischer.
- Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol.* 27, 1411–1423. doi: 10.1037/0278-7393.27.6.1411
- Graf Estes, K., Evans, J. L., and Else-Quest, N. (2007). Differences in non-word repetition performance for children with and without specific language impairment: a meta-analysis. *J. Speech Lang. Hear. Res.* 50, 177–195. doi: 10.1044/1092-4388(2007/015)
- Grimm, A., Ferré, S., 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*, (Amsterdam: IASCL).
- 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
- 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).
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459
- Grosjean, F. (2016). “The complementarity principle and its impact on processing, acquisition and dominance,” in *Language Dominance in Bilinguals*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 66–85.
- 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 (Philadelphia, PA: 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., and Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Front. Commun.* 2:19. doi: 10.3389/fcomm.2017.00016
- 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 (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, Kraków*.
- 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 Grammatica Generativa* 36, 146–181.
- Hamann, C., Tuller, L., Monjauze, C., Delage, H., and Henry, C. (2007). “(Un)successful subordination in French-speaking children and adolescents with SLI,” in *Proceedings of BUCLD 31*, eds H. Cauntulton, S. Kalatilake, and I. Woo (Somerville, MA: Cascadia Press), 286–297.
- 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., 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
- Hopkins, B., and Skellam, J. G. (1954). A new method for determining the type of distribution of plant individuals. *Ann. Bot.* 18, 213–227. doi: 10.1093/oxfordjournals.aob.a083391
- Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933
- IBM SPSS 24 (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.
- International Association of Logopedics and Phoniatrics [IALP] (2011). *Recommendations for Working With Bilingual Children*. Available at: [http://www.specchioriflesso.net/media/162083/linea\\_guida\\_bilingui\\_ialp\\_may\\_2011.pdf](http://www.specchioriflesso.net/media/162083/linea_guida_bilingui_ialp_may_2011.pdf)
- Jakubowicz, C. (2005). The language faculty: (Ab)normal development and interface constraints. *Paper Presented at Generative Approaches to Language Acquisition, GALA, 2005*, Siena.
- Jakubowicz, C. (2011). Measuring derivational complexity: new evidence from typically developing and SLI learners of L1 French. *Lingua* 121, 339–351. doi: 10.1016/j.lingua.2010.10.006
- 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: 10.1080/02699206.2018.1437222
- Jakubowicz, C., and Tuller, L. (2008). “Specific language impairment in French,” in *Studies in French applied linguistics*, ed. D. Ayoun (Amsterdam: John Benjamins), 97–134.

- 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)
- Kashef, R., and Kamel, M. S. (2008). “Efficient bisecting k-medoids and its application in gene expression analysis,” in *ICIAR 2008 Proceedings of the 5th International Conference on Image Analysis and Recognition, LNCS 5112*, eds A. Campilho and M. Kamel (Heidelberg: Springer), 423–434. doi: 10.1007/978-3-540-69812-8\_42
- Kassambara, A., and Mundt, F. (2017). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses, R Package Version 1.0.5*. Available at: <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf> [accessed June 1, 2018].
- Kaufman, L., and Rousseeuw, P. J. (1987). “Clustering by means of Medoids,” in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, ed. Y. Dodge (Basel: Birkhäuser Basel), 405–416.
- Kaufman, L., and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.
- Kohnert, K. (2010). Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. *J. Commun. Disord.* 43, 456–473. doi: 10.1016/j.jcomdis.2010.02.002
- 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: 10.1016/j.jcomdis.2010.02.002
- Kosmidis, I. (2018). *brglm2: Bias Reduction in Generalized Linear Models. R Package Version 0.1.7*. Available at: <https://github.com/ikosmidis/brglm2>
- Krishna, S., Yesu Babu, A., and Kiran Kumar, R. (2018). “Determination of optimal clusters for a Non-hierarchical clustering paradigm K-Means algorithm,” in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, eds N. Devarakonda, N. Chaki, and A. Cortesi (Singapore: Springer), 301–316. doi: 10.1007/978-981-10-6319-0\_26
- Kupisch, T. (2008). “Dominance, mixing and cross-linguistic influence: on their relation in bilingual development,” in *First Language Acquisition of Morphology and Syntax: Perspectives Across Languages and Learners*, eds P. Guijarro-Fuentes, P. Larrañaga, and J. Clibbens (Amsterdam: John Benjamins), 209–234.
- Kutner, M. H., Nachtsheim, C. J., and Neter, J. (2004). *Applied Linear Regression Models*, 4th Edn. Chicago, IL: McGraw-Hill/ Irwin.
- Leclercq, A. L., Maillart, C., and Majerus, S. (2013). Nonword repetition problems in children with specific language impairment: a deficit in accessing long-term linguistic representations? *Top. Lang. Disord.* 33, 238–254. doi: 10.1097/TLD.0b013e31829dd8c9
- 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 *Proceedings of the 2015 GASLA*, eds D. Stringer, J. Garrett, B. Halloran, and S. Mossman (Somerville, MA: Cascadia Proceedings), 124–135.
- Leonard, L. (1998). *Children With Specific Language Impairment*. Cambridge: MIT Press.
- Leonard, L. (2007). Processing limitations and the grammatical profile of children with specific language impairment. *Adv. Child Dev. Behav.* 35, 139–171. doi: 10.1016/B978-0-12-009735-7.50009-8
- Leonard, L. (2014). *Children With Specific Language Impairment*, 2nd Edn. Cambridge: MIT Press.
- Leonard, L. (2014). Children with specific language impairment and their contribution to the study of language development. *J. Child Lang.* 41, 37–48. doi: 10.1017/S0305000914000130
- Levshina, N. (2015). *How to do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/z.195
- Lindener, K. (2002). Finiteness and children with specific language impairment: an exploratory study. *Linguistics* 40, 797–847. doi: 10.1515/ling.2002.031
- Long, S. L. (1997). *Regression Models for Categorical and Limited Dependent Variables*. London: SAGE Publications.
- Maddieson, I. (2006). Correlating phonological complexity: data and validation. *Linguist. Typol.* 10, 106–123.
- Maddieson, I., Flavier, S., Marsico, E., and Pellegrino, F. (2011). *LAPSYd: Lyon-Albuquerque Phonological Systems Databases, Version 1.0*. Available at: <http://www.lapsyd.dcl.ish-lyon.cnrs.fr/>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2017). *Cluster: Cluster Analysis Basics and Extensions. R package version 2.0.1*.2015.
- Marinis, T. (2011). On the nature and cause of specific language impairment: a view from sentence processing and infant research. *Lingua* 121, 463–475. doi: 10.1016/j.lingua.2010.10.010
- 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*, eds J. Rothman and S. Unsworth (Amsterdam: John Benjamins), 265–276.
- Marinis, T., and van der Lely, H. (2007). On-line processing of wh-questions in children with G-SLI and typically developing children. *Int. J. Lang. Commun. Disord.* 42, 557–582. doi: 10.1080/13682820601058190
- Marshall, C., Ebbels, S., Harris, J., and van der Lely, H. K. (2002). Investigating the impact of prosodic complexity on the speech of children with specific language impairment. *UCL Work. Pap. Linguist.* 14, 43–66.
- Marshall, C., Harris, J., and van der Lely, H. (2003). “The nature of phonological representations in children with grammatical specific language impairment,” in *Proceedings of the University of Cambridge First Postgraduate Conference in Language Research*, (Cambridge: Institute of Language Research Cambridge), 511–517.
- Marshall, C. R. (2014). Word production errors in children with developmental language impairments. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20120389. doi: 10.1098/rstb.2012.0389
- 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
- Meir, N. (2018). Morpho-syntactic abilities of unbalanced bilingual children: a closer look at the weaker language. *Front. Psychol.* 9:1318. doi: 10.3389/fpsyg.2018.01318
- Meir, N., and Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic Status (SES) and bilingualism on children’s vocabulary and verbal short-term memory. *Front. Psychol.* 8:1442. doi: 10.3389/fpsyg.2017.01442
- Meir, N., Walters, J., and Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: the impact of L1 and L2 properties. *Int. J. Biling.* 20, 421–452. doi: 10.1177/1367006915609240
- Meir, N., Walters, J., and Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguist. Approaches Biling.* 7, 514–553. doi: 10.1075/lab.15007.mei
- Meisel, J. (2007). The weaker language in early child bilingualism: acquiring a first language as a second language? *Appl. Psycholinguist.* 28, 495–514. doi: 10.1017/S0142716407070270
- Meisel, J. (2009). Second language acquisition in early childhood. *Zeitschrift Sprachwissenschaft* 28, 5–34. doi: 10.1515/ZFSW.2009.002
- Messer, M. H., Leseman, P. P. M., Boom, J., and Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *J. Exp. Child Psychol.* 105, 306–323. doi: 10.1016/j.jecp.2009.12.006
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism. Re-examining the Age Factor*. Amsterdam: Benjamins. doi: 10.1075/sibil.39
- Novogrodsky, R., and Kreiser, V. (2015). What can errors tell us about specific language impairment deficits? Semantic and morphological cuing in a sentence completion task. *Clin. Linguist. Phon.* 29, 812–825. doi: 10.3109/02699206.2015.1051239
- Oetting, J. B., Cleveland, L., and Cope, R. (2008). Empirically derived combinations of tools and clinical cutoffs: an illustrative case with a sample of culturally/linguistically diverse children. *Lang. Speech Hear. Serv. Schl.* 39, 44–53. doi: 10.1044/0161-1461(2008/005)
- Orsolini, M., Sechi, E., Maronato, C., Bonvino, E., and Corcelli, A. (2001). Nature of phonological delay in children with specific language impairment. *Int. J. Lang. Commun. Disord.* 36, 63–90. doi: 10.1080/13682820150217572

- Ott, S., van de Vijver, R., and Höhle, B. (2006). The effect of phonotactic constraints in German-speaking children with delayed phonological acquisition: evidence from production of word-initial consonant clusters. *Adv. Speech Lang. Pathol.* 8, 323–334. doi: 10.1080/14417040600970622
- Paradis, J. (2009). *Oral Language Development in French and English and the Role of Home Input Factors*. Conseil scolaire centre-nord: Edmonton.
- 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.
- Paradis, J. (2011). Individual differences in child English second-language acquisition: comparing child-internal and child-external factors. *Linguist. Approaches Biling.* 1, 834–847. doi: 10.1044/jslhr.4304.834
- Paradis, J., Crago, M., 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., Emmert, 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., and Genesee, F. (1996). Syntactic acquisition in bilingual children: autonomous or interdependent? *Stud. Second Lang. Acquis.* 18, 1–25. doi: 10.1017/S0272263100014662
- Paradis, J., Genesee, F., and Crago, M. (2011a). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*, 2nd Edn. Baltimore, MD: Brookes.
- Paradis, J., Nicoladis, E., Crago, M., and Genesee, F. (2011b). Bilingual children’s acquisition of the past tense: a usage-based approach. *J. Child Lang.* 38, 544–578. doi: 10.1017/S0305000910000218
- Paradis, J., and Jia, R. (2016). 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.* 20:1. doi: 10.1111/desc.12433
- Paradis, J., Tulpar, 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
- Patel, A., and Singh, P. (2013). New approach for K-mean and K-medoids algorithm. *Int. J. Comput. Appl. Technol. Res.* 2, 1–5.
- Pearson, B. Z., Fernald, S. C., Lewedag, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants (ages 10 to 30 months). *Appl. Psycholinguist.* 18, 41–58. doi: 10.1017/S0142716400009863
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., and Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* 49, 1373–1379. doi: 10.1016/S0895-4356(96)00236-3
- Petersen, J. (1988). Word-internal code-switching constraints in a bilingual child’s grammar. *Linguistics* 26, 479–493. doi: 10.1515/ling.1988.26.3.479
- 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
- Polišenská, K., Chiat, S., and Roy, P. (2015). Sentence repetition: what does the task measure? *Int. J. Lang. Commun. Disord.* 50, 106–118. doi: 10.1111/1460-6984.12126
- Rainey, C. (2016). Dealing with separation in logistic regression models. *Polit. Anal.* 24, 339–355. doi: 10.1093/pan/mpw014
- Reilly, S., Wake, M., Ukoumunne, O., Bavin, E., Prior, M., Cini, E., et al. (2010). Predicting language outcomes at 4 years of age: findings from early language in victoria study. *Pediatrics* 126, 1530–1537. doi: 10.1542/peds.2010-0254
- Rice, M., Noll, K., and Grimm, H. (1997). An extended optional infinitive stage in German-speaking children with specific language impairment. *Lang. Acquis.* 61, 255–295. doi: 10.1207/s15327817la0604\_1
- Rice, M., and Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *J. Speech Lang. Hear. Res.* 39, 1239–1257. doi: 10.1044/jshr.3906.1239
- Rispens, J., and Baker, A. (2012). Nonword repetition: the relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *J. Speech Lang. Hear. Res.* 55, 683–694. doi: 10.1044/1092-4388(2011/10-0263)
- 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.
- Rizzi, L. (2013). Locality. *Lingua* 130, 169–186. doi: 10.1016/j.lingua.2012.12.002
- Roeper, T. (2004). Diagnosing language variations: underlying principles for syntactic assessment. *Semin. Speech Lang.* 25, 41–56. doi: 10.1055/s-2004-824825
- Roeper, T. (2011). The acquisition of recursion: how formalism articulates in the child’s path. *Biolinguistics* 5, 57–87.
- 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. Cogn.* 15, 39–57.
- Roy, P., and Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: evidence from typically developing children. *J. Speech Lang. Hear. Res.* 47, 223–234. doi: 10.1044/1092-4388(2004/019)
- Roy, P., Chiat, S., and Dodd, B. (2014). *Language and Socioeconomic Disadvantage: From Research to Practice*. London: University of London.
- Royal College of Speech and Language Therapists Specific Interest Group in Bilingualism [RCSLT] (2007). *Good Practice for Speech and Language Therapists Working With Clients From Linguistic Minority Communities*. London: RCSLT.
- R-Studio (2012). *RStudio: Integrated Development Environment for R (Version 0.99.892)*. Boston, MA: R-Studio.
- Scheidenes, M., and Tuller, L. (2018). Using clausal embedding to identify language impairment in sequential bilinguals. *Biling. Lang. Cogn.* 1–19. doi: 10.1017/S1366728918000949
- Schroeder, C., and Dollnick, M. (2013). “Mehrsprachige gymnasien mit türkischem hintergrund schreiben auf türkisch,” in *Mehrsprachig in Wissenschaft und Gesellschaft. Tagungsband*, eds C. Riemer, H. Brandl, E. Arslan, and E. Langelahn (Bielefeld: University of Bielefeld), 101–114.
- 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: Hofgrefe Verlag.
- Soni, K. G., and Patel, A. (2017). Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. *Int. J. Comput. Intell. Res.* 13, 899–906.
- Sua-Kay, E., and Santos, M. E. (2014). *Grelha de Avaliacao da Linguagem – Nível Escolar (GOL-E) 2nd edição revista*. Alcoitao: Escola Superior de Saúde do Alcoitao.
- Tagliamonte, S. A., and Baayen, R. H. (2012). Models, forests, and tress of York English: was/ were variation as a case study for statistical practice. *Lang. Var. Change* 24, 135–178. doi: 10.1017/S0954394512000129
- Tamburelli, M., and Jones, G. (2013). Investigating the relationship between nonword repetition and syllable structure in typical and atypical language development. *J. Speech Lang. Hear. Res.* 56, 708–720. doi: 10.1044/1092-4388(2012/11-0171)
- Theodorou, E., Kambanaros, M., and Grohmann, K. K. (2017). Sentence repetition as a tool for screening morphosyntactic abilities of bilingual children with SLI. *Front. Psychol.* 8:2104. doi: 10.3389/fpsyg.2017.02104
- Thordardottir, E. (2012). *Proposed Diagnostic Procedures and Criteria for Bi-SLI COST Action Studies*. 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. doi: 10.21832/9781783093137-014
- 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
- Tibshirani, R., Guenther, W., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63, 411–423. doi: 10.1111/1467-9868.00293
- Tomblin, B. J., Records, N. L., and Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 39, 1284–1294. doi: 10.1044/jshr.3906.1284
- Topbaş, S., and Güven, S. (2013). *Türkçe Erken Dil Gelism Testi (TEDYL) (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.
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prévost, P., et al. (2018). Identifying language impairment in bilingual children in France and in Germany. *Int. J. Lang. Commun. Disord.* 53, 888–904. doi: 10.1111/1460-6984.12397
- Unsworth, S. (2016). “Amount of exposure as a proxy for dominance in bilingual language acquisition,” in *Language Dominance in Bilinguals*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173.
- Unsworth, S., Chondrogianni, V., and Skzarabela, B. (2018). Experiential measures can be used as a proxy for language dominance in bilingual language acquisition research. *Front. Psychol.* 9:1809. doi: 10.3389/fpsyg.2018.01809
- van der Lely, H. (1998). SLI in children: movement, economy, and deficits in the computational-syntactic system. *Lang. Acquis.* 7, 161–192. doi: 10.1207/s15327817la0702-4\_4
- van der Lely, H. (2005). Domain-specific cognitive systems: insight from grammatical specific language impairment. *Trends Cogn. Sci.* 9, 53–59. doi: 10.1016/j.tics.2004.12.002
- Vinther, T. (2002). Elicited imitation: a brief overview. *Int. J. Appl. Linguist.* 12, 54–73. doi: 10.1111/1473-4192.00024
- 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)
- Yip, V., and Matthews, S. (2006). Assessing language dominance in bilingual acquisition: a case for mean length utterance differentials. *Lang. Assess. Q.* 3, 97–116. doi: 10.1207/s15434311laq0302\_2
- Zebib, R., Henri, G., Khomsi, A., Messara, C., and Hreich, E. (2017). *Batterie d'Evaluation du Langage Oral chez l'enfant libanais (ELO-L)*. Kerserwan: LTE.
- Zurer-Pearson, B. (2010). We can no longer afford a monolingual norm. *Appl. Psycholinguist.* 31, 339–343. doi: 10.1017/S014271640999052X

**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 AG and handling Editor declared their shared affiliation.

Copyright © 2019 Abed Ibrahim and Fekete. 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.



# APPENDIX

**TABLE A1** | Correlations between background variables for BITD and BiSLI groups collapsed.

			Age	Positive_Earl_Dvt	AcO	SES	LDI	LoE
Spearman-Rhc	Age	Corr. coeff.	1.000	0.032	0.207	0.237	0.086	0.445**
		sig. (2-tailed)		0.816	0.126	0.078	0.514	0.001
		N	56	56	56	56	56	56
	Positive_Early_Dvt	Corr. coeff.	0.032	1.000	0.103	0.064	−0.057	−0.146
		sig. (2-tailed)	0.816		0.450	0.639	0.676	0.282
		N	56	56	56	56	56	56
	AoO	Corr. coeff.	0.207	0.103	1.000	−0.051	−0.351**	−0.670***
		sig. (2-tailed)	0.126	0.450		0.711	0.008	0.000
		N	56	56	56	56	56	56
	SES	Corr. coeff.	0.237	0.064	−0.051	1.000	0.269*	0.231
		sig. (2-tailed)	0.078	0.638	0.711		0.045	0.087
		N	56	56	56	56	56	56
	LDI	Corr. coeff.	0.089	−0.057	−0.351**	0.269*	1.000	0.428**
		sig. (2-tailed)	0.514	0.679	0.008	0.045		0.001
		N	56	56	56	56	56	56
	LoE	Corr. coeff.	0.44**	−0.146	−0.670***	0.231	0.428**	1.000
		sig. (2-tailed)	0.001	0.282	0.000	0.087	0.001	
		N	56	56	56	56	56	56

\*The correlation is significant at the level of 0.05 (2-tailed).

\*\*The correlation is significant at the level of 0.01 (2-tailed).

\*\*\*The correlation is significant at the level of 0.001 (2-tailed).



# Vocabulary, Metalinguistic Awareness and Language Dominance Among Bilingual Preschool Children

**Carmit Altman<sup>1\*</sup>, Tamara Goldstein<sup>1</sup> and Sharon Armon-Lotem<sup>2</sup>**

<sup>1</sup> School of Education at Bar Ilan University, Ramat Gan, Israel, <sup>2</sup> Department of English Literature and Linguistic, Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel

## OPEN ACCESS

### Edited by:

Cornelia Hamann,  
University of Oldenburg, Germany

### Reviewed by:

Marta Marecka,  
Jagiellonian University, Poland  
Katja Francesca Cantone,  
Universität Duisburg-Essen, Germany

### \*Correspondence:

Carmit Altman  
carmit.altman@biu.ac.il

### Specialty section:

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

**Received:** 29 April 2018

**Accepted:** 21 September 2018

**Published:** 23 October 2018

### Citation:

Altman C, Goldstein T and  
Armon-Lotem S (2018) Vocabulary,  
Metalinguistic Awareness  
and Language Dominance Among  
Bilingual Preschool Children.  
Front. Psychol. 9:1953.  
doi: 10.3389/fpsyg.2018.01953

Awareness of language structure has been studied in bilinguals, but there is limited research on how language dominance is related to metalinguistic awareness, and whether metalinguistic awareness predicts vocabulary size. The present study aims to explore the role of language dominance in the relation between vocabulary size in both languages of bilingual children and metalinguistic awareness in the societal language. It evaluates the impact of two metalinguistic awareness abilities, morphological and lexical awareness, on receptive and expressive vocabulary size. This is of special interest since most studies focus on the impact of exposure on vocabulary size but very few explore the impact of the interaction between metalinguistic awareness and dominance. 5–6-year-old preschool children with typical language development participated in the study: 15 Russian-Hebrew bilingual children dominant in the societal language (SL) Hebrew, 21 Russian-Hebrew bilingual children dominant in the Heritage language (HL) Russian and 32 monolingual children. Dominance was determined by relative proficiency, based on standardized tests in the two languages. Tasks of morphological and lexical awareness were administered in SL-Hebrew, along with measures of receptive and expressive vocabulary size in both languages. Vocabulary size in SL-Hebrew was significantly higher for SL-dominant bilinguals (who performed like monolinguals) than for HL-dominant bilinguals, while HL-Russian vocabulary size was higher for HL-dominant bilinguals than for SL-dominant bilinguals. A hierarchical regression analyzing the relationship between vocabulary size and metalinguistic awareness showed that dominance, lexical metalinguistic awareness and the interaction between the two were predictors of both receptive and expressive vocabulary size. Morphological metalinguistic awareness was not a predictor of vocabulary size. The relationship between lexical awareness and SL-vocabulary size was limited to the HL-dominant group. HL-dominant bilinguals relied on lexical metalinguistic awareness, measured by fast mapping abilities, that is, the abilities to acquire new words, in expanding their vocabulary size, whereas SL-dominant bilinguals and monolinguals did not. This difference reflects the milestones of lexical acquisition the different groups have reached. These findings show that metalinguistic awareness should also be taken into consideration when evaluating the variables that influence vocabulary size among bilinguals through different ways in different dominance groups.

**Keywords:** bilingualism, Russian-Hebrew, metalinguistic awareness, dominance, vocabulary size

## INTRODUCTION

Language dominance among bilingual children can be defined by their relative proficiency in each language, but there is limited research on how language dominance is related to metalinguistic awareness, and whether metalinguistic awareness predicts vocabulary size. The present study aims to explore the role of language dominance in the relationship between vocabulary size in both languages of bilingual children and metalinguistic awareness in the societal language (SL). To achieve this aim, receptive and expressive vocabulary size is tested in both languages of Russian-Hebrew bilingual preschool children who are dominant in one of their languages. This is complemented by measuring metalinguistic awareness in the SL, Hebrew, and by analyzing the relations between vocabulary size and metalinguistic awareness.

### Vocabulary of Monolingual and Bilingual Children

Studies show that bilingual children score below monolingual age appropriate norms when vocabulary size is assessed in only one of their languages (Bialystok et al., 2010; Hoff et al., 2012; Spaulding et al., 2013). For example, Spanish-English bilingual students lag behind monolingual age matched peers in oral language abilities in SL English and in the heritage language (HL) Spanish (Tabors et al., 2003; Pérez et al., 2007; Uccelli and Pérez, 2007). In particular, English vocabulary skills were limited for children at 4 years of age (Pérez et al., 2007), with low levels of vocabulary and gaps between monolingual norms and bilingual children's scores persisting through first grade (Pérez and Rinaldi, 2006). When it comes to vocabulary size in bilinguals' HL, some studies show poor performance in both receptive and expressive vocabulary (Pearson et al., 1997; Uccelli and Pérez, 2007; Bialystok et al., 2010; Verhoeven et al., 2011; O'Toole et al., 2017), while there are other studies that do not show this effect (Umbel and Ki Oller, 1994; Winsler et al., 1999). Moreover, previous findings are not always consistent as to whether a receptive and expressive vocabulary gap (Keller et al., 2015) exists in both languages and if so which factors contribute to its existence. Umbel and Ki Oller (1994), for example, found that Spanish-English bilinguals in first, third, and sixth grade functioned comparably well on the HL Spanish receptive vocabulary test, while SL English receptive vocabulary performance increased with grade level. Furthermore, a receptive-expressive gap was found in a study of 124 Spanish-English bilingual children and 110 monolingual children (mean age = 5;7), for both groups, with a more robust gap amongst the bilinguals, in both languages (Gibson et al., 2012).

These inconsistent results might stem from different factors influencing whether bilingual children perform well or poorly on vocabulary size tests. Therefore, it is important to examine these factors. One often studied factor is exposure. Differences in vocabulary size between bilingual children have often been attributed to variations in the frequency of exposure (Pearson et al., 1997) and, sometimes, to variations in the context of exposure (Bialystok et al., 2010). The vocabulary gap between bilinguals and their monolingual peers is not surprising as

children exposed to two languages are likely to hear less of each language during the day than children who are exposed to only one language. Moreover, some words occur in contexts where only one of the languages is used (Fromkin et al., 2007). Consequently, by looking both at English receptive and expressive vocabulary of Spanish-English bilingual children, aged 5–7, Gross et al. (2014) found that bilinguals scored significantly below monolingual children on standardized measures, with bilinguals exposed to SL later lagging behind their peers who were exposed to SL earlier. However, when tested in both languages, the difference in cumulative expressive vocabulary size was no longer significant.

Yet another, less investigated factor is metalinguistic awareness, which might be mediated by language dominance. Metalinguistic awareness builds on earlier linguistic knowledge, which might vary by language dominance, across the two languages of a bilingual child. It is the aim of this paper to assess bilingual dominance and metalinguistic awareness as possible factors that may explain the contradictory results in the literature. The difference between bilinguals and monolinguals in vocabulary size and the gap between expressive and receptive vocabulary further highlight the importance of testing different dominance groups in order to understand the contribution of the relative proficiency in each language in each modality, expressive or receptive (Spaulding et al., 2013).

### Language Dominance Among Bilinguals

The term “language dominance” is used in the literature either for describing the relative proficiency of a bilingual person in the two languages (Gathercole and Thomas, 2009), or for the language the bilingual speaker has been mostly exposed to (Grosjean, 2008). One of the dilemmas which both researchers and language therapists face is how to define dominance (Yip and Matthews, 2007). A most common way is to examine a sample of the child's productions using one or more performance-based measures and to establish in this way the child's relative proficiency in his or her two languages. Following Unsworth (2015), this is the way language dominance is defined in the current study. Later age of onset of bilingualism is frequently associated with relative proficiency and more advanced HL outcomes (Hammer et al., 2012; Meir et al., 2016). Yet, age of onset of bilingualism is not necessarily an indicator of dominance, as simultaneous and sequential bilinguals may be found in both the HL-dominant and the SL-dominant groups (Foroodi Nejad and Paradis, 2009). Therefore, the bilingual children in the present study will not be divided into simultaneous and sequential bilinguals, but rather into two dominance groups by their relatively more proficient language.

Language proficiency of bilinguals is often associated with the extent to which vocabulary size in one or both languages meets the norms set for age matched monolinguals (Bialystok et al., 2010). However, bilingual children's performance may be more varied than monolingual performance as a result of the diversity in their language learning experience (Armon-Lotem et al., 2015). This variation in bilinguals' performance, often captured in terms of language dominance, might differ as a function of the language skill assessed, resulting in asymmetric linguistic development

(Montrul, 2016). While awareness of the formal structure of language has already been studied among bilinguals (Reder et al., 2013), relatively little is known about the association between metalinguistic awareness and vocabulary size in the context of bilingual dominance. It is the aim of this paper to shed light on the relationship between receptive and expressive vocabulary of nouns and verbs in both languages and metalinguistic awareness (morphological and lexical) in the SL among bilinguals, who are dominant in one of their languages.

## Metalinguistic Awareness

Metalinguistic awareness is defined as the ability to distance oneself from the content of speech in order to reflect upon and manipulate the structure of language (Ramirez et al., 2013). Metalinguistic awareness requires the speaker to focus on the structure and form of the language and develops in later stages of language acquisition around the age of 5–6, building on earlier linguistic knowledge (Duncan et al., 2009). Metalinguistic awareness is a set of multiple skills (Bialystok et al., 2014) that are related to the formal aspects of language: phonological, morphological, syntactic and lexical awareness.

Some studies found a statistically significant difference between monolingual and bilingual children on metalinguistic awareness (e.g., Bialystok et al., 2005; Goldstein et al., 2005), pointing out that different skills and tasks might yield different results. For example, Reder et al. (2013) compared 52 French monolingual and 43 French-German bilingual children in first grade, on different metalinguistic skills. While bilingual children outperformed their monolingual peers in morphological compounds and syntactic awareness tasks, no differences were found in morphological affixes and phonological awareness tasks. They argued that due to the phonological similarities between the two languages (French and German), the bilingual children were not required to observe and compare the different linguistic aspects of each language (McBride-Chang et al., 2005). Yet, other studies have shown that bilingual speakers outperform monolingual speakers in metalinguistic awareness tasks (for review see Bialystok, 2001). In particular, in a meta-analysis of 63 studies consisting of 6,022 participants, Adesope et al. (2010) examined the cognitive correlates of bilingualism and found that bilingualism is related to enhanced metalinguistic awareness. The bilingual enhancement observed in the meta-analysis shows the importance of going beyond single studies, which in themselves do not show this effect. However, none of these studies examined the impact of dominance among bilinguals on metalinguistic awareness tasks as the present study intends to do with Russian, the HL, and Hebrew, the SL.

## Metalinguistic Abilities and Vocabulary Size

Vocabulary size is a major factor in language acquisition and as such, it is closely related to metalinguistic skills. On the one hand, vocabulary size is enhanced by metalinguistic abilities and on the other hand, metalinguistic abilities often benefit from a richer vocabulary. Yet, research investigating the metalinguistic abilities in bilinguals focus primarily on phonological awareness

and its contribution to reading skills (see example: Carlisle et al., 1999; Ibrahim et al., 2007). Some studies have indeed investigated phonological awareness and vocabulary in bilinguals showing a relationship between phonological awareness and vocabulary (Farnia and Geva, 2011). Children with poorer phonological awareness learned novel and non-novel words less accurately or more slowly (Hu and Schuele, 2005; Hu, 2008). Longitudinally, phonological awareness plays a role when words are relearned (Hu, 2003) and phonological processing of novel words is based on sublexical representations, which are phonological and unstructured (Marecka et al., 2018).

In comparison, there are hardly any similar studies for morphological and lexical awareness and their association to vocabulary size (Bowey, 1986; Reder et al., 2013). Morphological awareness relates to the ability to manipulate and reflect on morphological units within words (Cheung et al., 2010). It includes the explicit knowledge of the way in which words are built up by combining smaller meaningful units, such as roots, prefixes and suffixes (Guo et al., 2011). Studies have shown that morphological awareness can facilitate word recognition, learning of new words and reading comprehension (Chen et al., 2009; Kraut, 2015).

The importance of morphological awareness for vocabulary learning is well documented in monolingual children (Chen et al., 2012). Nagy et al. (2003) found a strong tie between vocabulary knowledge and morphological awareness, while McBride-Chang et al. (2005) showed that morphological structure awareness and morpheme identification predicted 10% of the variance in vocabulary size. These results underline the importance of examining the impact of different metalinguistic abilities on vocabulary separately in order to understand the variability in vocabulary size (Kuo and Anderson, 2006). Yet, to the best of our knowledge, very little is known about these connections in bilingual contexts in which children acquire vocabulary in two languages and the process might be at a different stage in each language.

Another form of metalinguistic awareness is lexical awareness, which includes conscious consideration of and the ability to manipulate different aspects of lexical competence (Nation, 2008). According to Aşık et al. (2015), lexical competence includes vocabulary size, depth and lexical organization. Nation (2008) argues that lexical awareness can help language learners increase their understanding of the different ways in which vocabulary is used, thus leading, for example, to growth in vocabulary size.

An easy way to measure lexical awareness is fast mapping. Fast mapping refers to the ability of a child to identify the meaning of a novel word after a limited number of exposures (Carey and Bartlett, 1978). It has been observed that growth in vocabulary size is related to fast mapping skills both in the initial stages of word learning (Behrend et al., 2001) and for later acquisition by older children (Braisby et al., 2001). Significant correlations were found between fast mapping performance and vocabulary size scores in early vocabulary acquisition (Kan and Kohnert, 2005; Gray, 2006; Kan et al., 2014), and for older children (ages 4;6–7) with expressive vocabulary scores (Braisby et al., 2001).



Within the developmental lexical principles framework (DLPF) (Golinkoff et al., 1994; Mervis and Bertrand, 1994), fast mapping involves six principles that govern vocabulary acquisition and apply to all languages. The first three include the understanding that words (a) have a reference in the world, (b) can extend to similar referents, and (c) refer to whole objects rather than their parts. These three principles are operative at the onset of lexical acquisition and help in acquiring early vocabulary. The principles which are more related to fast mapping are operative beyond early childhood, in older children and adults (Golinkoff et al., 1992), and are utilized in consciously monitoring the learning of novel words (Ramachandra et al., 2010). These three principles require the: (d) awareness of basic categories for generalization, (e) awareness of constraints on mapping novel names to nameless objects to meet mutual exclusivity, and (f) consideration of the use of conventional names for referents. Fast mapping is an appropriate measure of lexical awareness because the growth in vocabulary size benefits from the latter three principles that operate together. Previous studies have shown the relationship between lexical awareness (measured with this task) and vocabulary size in monolinguals (Behrend et al., 2001; Braisby et al., 2001). Bilinguals also need to apply such constraints when they map novel names to nameless objects. Yet bilinguals also need to learn two labels for the same object, one in each language. In order to abide by the above principles, they should be aware of the differences between the two vocabularies and of translation equivalents. Currently, little is known about the possible interaction between fast mapping and vocabulary size in the case of bilingual children. Of the very few studies of fast mapping and vocabulary size among bilingual children, Kan and Kohnert (2008) do not find such an interaction.

Kan and Kohnert (2008) tested lexical awareness (via fast mapping) and vocabulary size in both the HL (Hmong) and the SL (English) of sequential bilingual children with typical language development (TLD), aged 3–5. In contrast to previous findings with monolingual children, the researchers found that the bilingual children's fast mapping performance was not related to age or existing vocabulary size in either language. On the other hand, there were significant correlations between vocabulary size and fast mapping across the two languages. For example, fast mapping in English (SL) was negatively correlated with vocabulary size in Hmong (HL), with lower fast mapping abilities in English for children who had larger vocabulary size in Hmong.

According to Kan and Kohnert (2008), this cross-linguistic relation suggests that fast mapping in the SL of bilingual children is not a direct measure of vocabulary size in that language, in contrast to what has been observed in monolingual children. There is, however, a cross-linguistic relationship between fast mapping and vocabulary size in sequential bilinguals – vocabulary size has a negative impact on fast mapping skills in the other language. While the authors made no direct reference to dominance, they suggested that a difference in vocabulary size in either of the languages can perhaps reflect a different stage of language development of sequential bilinguals when compared to monolinguals. Since dominance might be important, but was not considered in this study, we want to replicate the design with participants who are grouped by dominance.

To conclude, although researchers have examined the individual contributions of different metalinguistic abilities to the bilingual lexicon, very few have examined morphological and lexical metalinguistic awareness simultaneously, and even less so with regard to vocabulary size in both languages (McBride-Chang et al., 2005) among bilinguals differing in language dominance.

## Present Study

The present study aims to explore the impact of language dominance on the possible connections between vocabulary size in both languages and metalinguistic awareness in the SL. It is hypothesized that:

- (1) Dominance, measured by relative proficiency, will impact vocabulary size in both languages.
- (2) Fast mapping used to measure lexical awareness is language neutral and is important for lexical growth (Nation, 2001). It will show a stronger relation to vocabulary size at earlier stages in acquisition and by inference in the less dominant language.
- (3) Morphological metalinguistic awareness might be sensitive to language specific knowledge, which requires higher proficiency (Bialystok and Barac, 2012) in the target languages. Therefore, it will show stronger relations with vocabulary size in later stages of acquisition or in the more dominant language.
- (4) Fast mapping as a measure of lexical awareness, which is language neutral, is more likely to benefit receptive and expressive vocabulary size than morphological awareness, which is language specific.

In order to test these hypotheses, the study will first examine vocabulary size and metalinguistic awareness separately and then will turn to the relation between the two. Expressive and receptive vocabulary size, as well as morphological awareness and lexical awareness via a fast mapping task, will be tested among HL-dominant and SL-dominant bilingual children with TLD and their monolingual peers. The study is the first to investigate this relation among Russian-Hebrew bilinguals.

The relationships between the two metalinguistic awareness tasks and vocabulary in the context of bilingualism has only rarely been investigated (McBride-Chang et al., 2005). Based on research among monolingual children, correlations are to be expected between the two metalinguistic tasks (morphological and lexical) and vocabulary size in both languages, and in particular between lexical awareness and vocabulary which may be sensitive to dominance.

## MATERIALS AND METHODS

### Participants

Sixty-eight preschool children with TLD aged 58–78 months ( $M = 68.18$ ,  $SD = 4.66$ ) participated in the present study. Children with different language status formed three language groups: 15 SL-dominant children, 21 HL-dominant bilingual children, and 32 monolingual Hebrew children that served as reference for comparison. Children with hearing impairment, exposure to SL

for less than a year or parental concern regarding their child's language development were excluded from the study. Consent forms were sent to 136 children, out of which eighty were approved. After data was collected, 12 children were excluded from the study after scoring below monolingual and bilingual norms in the language proficiency tests. Inclusion of a bilingual child in the current study was based on a score at or above the provisional bilingual norm (Armon-Lotem and Meir, 2016) in at least one of their languages. Almost all of the participants were born in Israel except for one who was born outside the country and immigrated at the age of 1 year and 10 months. All children attended public preschools in Israel where the language of instruction is Hebrew. Age of onset of bilingualism was determined in months based on parent reports. All children scored above 85 in the "Raven Progressive Matrices" intelligence test (Raven, 1938).

In order to assess children's language performance in Hebrew, the *Goralnik Screening Test for Hebrew* (Goralnik, 1995) was used. The test includes six subtests: sentence repetition, comprehension, expression, pronunciation, vocabulary, and story-telling sub-tests. The scores are raw scores, with a total of 180 points. The Hebrew cut-off point conforms to former studies of bilingual children in Israel and has provisional bilingual norms (Iluz-Cohen and Armon-Lotem, 2013; Armon-Lotem, 2014; Altman et al., 2016). In order to assess the language performance of the bilingual children in their HL (Russian), the *Russian Language Proficiency Test for Multilingual Children* (Gagarina et al., 2010) was used. The task has a provisional bilingual norm for Russian-Hebrew bilinguals (with a cut-off point of -1.25 SD; Armon-Lotem and Meir, 2016). The raw scores in each screening test were normalized using the provisional norms.

For the present study, dominance was judged based on linguistic performance in two screening tests composed of several sub-tests (e.g., grammar, morphology) testing several domains in each language rather than focusing on a specific domain in order to reflect bilinguals' performance on a wide range of HL and SL skills. An index of relative proficiency based on the differences between the two language scores, following Cromdal (1999), was calculated and used to determine the bilinguals' dominance. Relative proficiency was calculated by deducing the normalized HL score from the normalized SL score. This resulted in negative scores for children whose HL scores were higher than their SL scores and positive scores for children whose HL scores were lower than their SL scores. Dominance was measured by a gap of

one standard deviation or more between the more proficient and less proficient language as measured by the language screening tests. The index was then used to separate the children into more dominant in the HL or more dominant in the SL. Children's demographic information appears in **Table 1**.

ANOVAs conducted to examine language proficiency differences between the bilingual dominance group showed differences in Hebrew  $F(1,34)=28.61$ ,  $p < 0.001$  and Russian proficiency  $F(1,34)=51.52$ ,  $p < 0.001$ . Additional ANOVAs show significant differences in terms of age of onset (AoO) as well as in length of exposure (LoE),  $F(1,34) = 17.95$ ,  $p < 0.001$  and  $F(1,35) = 20.45$ ,  $p < 0.001$ , respectively. These differences were expected since AoA and LoE are known to influence dominance. A one-way ANOVA investigating whether there are difference between the three language status groups showed a significant different  $F(2,65) = 27.4$ ,  $p < 0.001$ . A Bonferroni *post hoc* test yielded significant differences in Hebrew proficiency between monolinguals and HL-dominant bilinguals ( $p < 0.001$ ) and between SL-dominant bilinguals and HL-dominant peers ( $p < 0.001$ ) as expected due to the relative dominance in the languages, with no difference between the SL-dominant group and monolinguals. It should also be noted that no age differences were detected among the three groups  $F(2,65) = 0.66$ ,  $p > 0.05$ .

## Measures

### Cross Linguistic Lexical Task (CLT)

Children's vocabulary size in both languages was assessed with the Hebrew version of the LITMUS CLT-task (Haman et al., 2015; Altman et al., 2017; O'Toole et al., 2017), and the Russian version of the LITMUS CLT task<sup>1</sup> (Gagarina and Nenonen, 2017, Unpublished). Both versions of LITMUS CLT contain four separate subtests, measuring receptive, and expressive nouns and verbs separately. Receptive vocabulary is tested through a picture selection task with four pictures and expression through a naming task. Each subtest is composed of 32 items scored as correct or incorrect using the classification of responses described for LITMUS CLT (Haman et al., 2015). The final score is assigned to each subset as a percentage of correct responses out of 32.

### Morphological Awareness Task

A morphological awareness task was developed for Hebrew following McBride-Chang et al. (2005). The task included 14

<sup>1</sup>The overall reliability of this task is  $\alpha = 0.961$  for the Hebrew version and  $\alpha = 0.956$  for the Russian version.

**TABLE 1 |** Background and language proficiency information of participants.

Deographic variable	Monolinguals [N = 32]	SL-dominant bilinguals [N = 15]	HL-dominant bilinguals [N = 21]	df	F
Age in months	67 (4.3)	69.33 (4.67)	69.66 (4.75)	2,65	2.61
Hebrew proficiency (z-score)	0.19 (0.81)	0.46 (0.65)	-1.71 (1.47)	2,65	27.4***
Russian proficiency (z-score)	NA	-2.67 (1.79)	0.57 (0.88)	1,34	51.52***
Age of onset	NA	14.33 (19.28)	41.38 (18.59)	1,34	17.95***
Length of exposure	NA	54 (21.63)	28.66 (16.80)	1,34	15.64***

\*\*\* $p < 0.001$ .

test items that test consonantal root awareness (Hebrew being a Semitic language) and lexical compound awareness. For each item, the child is presented with two pictures of homophones that sound the same but have different meanings, and sometimes, different roots. The test items contained either two homophone verbs, two homophone nouns, or a homophone noun and verb. The examiner names each one orally. The child is then presented with the target item; a word or a lexical compound derived from one of the meanings of the homophone. The child is asked to choose the picture that corresponds best to the meaning of the target item. This requires knowledge that the words share the same root. The prompt in this task was: “Which of the pictures is more related to the word” . . . ? For example, the child is shown two pictures: “*or*” (light) and “*or*” (skin), and is asked to match correctly the word “*teura*” (lighting) to the target picture depicting “*or*” which shares the same root. A second example is “*yalda*” (a girl) and “*yalda*” (gave birth) – and the lexical compound “*erec moledet*” (place of birth). In this case, both pictures share the same root with the target item, but only one shares the meaning.

Each item includes an open question asking the child to explain his answer (“why did you choose this answer?”) in order to examine in a more qualitative manner the children’s responses and what they could reveal about their metalinguistic ability. A certain concern was raised that this task may tap into semantic association knowledge due to the use of pictures. Nevertheless, the pictures were considered necessary in order to administer and adapt this task to preschool children. The final score was assigned as a percentage of correct responses out of 14. The overall reliability of this task is  $\alpha = 0.54$ .

### Lexical Awareness Task

A fast mapping task was used to test lexical awareness (Kan and Kohnert, 2008). Novel bisyllabic non-words (CVCVC, e.g., *renil*, *tumof*, *pamig*, *xemog*) were presented to the children. The novel words were not easily associated with any existing referent in either language in order to minimize the possibility for phonological or semantic associations. A PowerPoint presentation was used to present children with an undersea creature who was teaching them the names of undersea objects. In the first stage, the child was simultaneously presented with four pictures on the screen and was asked to recognize a novel object among three known distractors (“Where is the *pamig*?”). The novel referent was presented among known objects to measure mutual exclusivity. After the child identified the object, she got a confirmation (*Right, this is the pamig*), or correction (*Are you sure? I think this is the pamig*), and was asked to repeat the word (*Can you say pamig?*). In total, the child was exposed to the word three times and was asked to repeat it once. In the next stage, the child was asked to identify the novel word with a referent that had the same shape but a different color among a second set of objects, two known and two novel. The child was asked again “Where is the *pamig*?” This measured receptive generalization skills, which are important since the child has to distinguish between the new word and other new concepts not known to him. This procedure was repeated four times with different items. To make it fun for the children a memory game followed in which the children were

asked to name all new objects. One point was assigned to each correct response. Due to a high correlation between the mutual exclusivity and the generalization measures ( $r = 0.753$ ,  $p < 0.001$ ), only the generalization measure, which is the closest indication of the child’s acquisition of the new word, was chosen to measure the child’s lexical awareness skill, yielding a maximum score of four.

### Procedure

The children were assessed individually in their preschool or in their homes in a private room for two sessions unless a specific child required more time. The children participated voluntarily and each child received a small reward (a sticker or a toy) at the end of each session as a token of appreciation to encourage their continuous collaboration. All responses were both audio-recorded and manually recorded on a response sheet. Parental consent was obtained, during which parents answered a short background questionnaire concerning demographic and language acquisition information, and the children’s oral assent was secured. The study was approved by the university IRB and by the Israeli Ministry of Education.

### Data Analysis

The information obtained from the four parts of the LITMUS CLT task in each language was calculated as a percentage of correct responses. The size of expressive and receptive vocabulary was calculated by combining the nouns and verbs and calculating the percentage of correct responses. The choice to present the results for both receptive and expressive vocabulary reflects the reported gap between the two, especially among bilingual children (Gibson et al., 2012), and the possibility that this gap is a reflection of the need to suppress the competition between the two languages in a naming task which could be sensitive to dominance. Consequently, a series of multivariate analyses of variance as well as ANOVAs were conducted to compare between bilingual dominance groups on HL and SL vocabulary size measures and between bilingual dominance groups and monolinguals on SL vocabulary size.

The metalinguistic awareness tasks were calculated separately as a percentage of correct responses (morphological and lexical). Relative proficiency was used as a measure of dominance for the hierarchical regression analyses. Following a comparison of the metalinguistic awareness measures across the bilingual dominance groups and the monolingual children, hierarchical regression was conducted introducing relative proficiency first, then the metalinguistic awareness tasks and finally the interactions between relative proficiency and metalinguistic awareness. The choice of hierarchical regression was motivated by the desire to explore the relative contribution of each predictor. The hierarchical regressions were conducted separately for receptive and expressive vocabulary in both the HL and the SL of all bilingual children as one group. As we used hierarchical regression with 5 predictors the model could be prone to overfitting. Thus, in order to confirm the results we further used linear regressions to test only the two metalinguistic predictors for each of the dominance groups separately as well as for the monolinguals, allowing us to tease apart their relative contribution to vocabulary size.

RESULTS

Vocabulary Measures

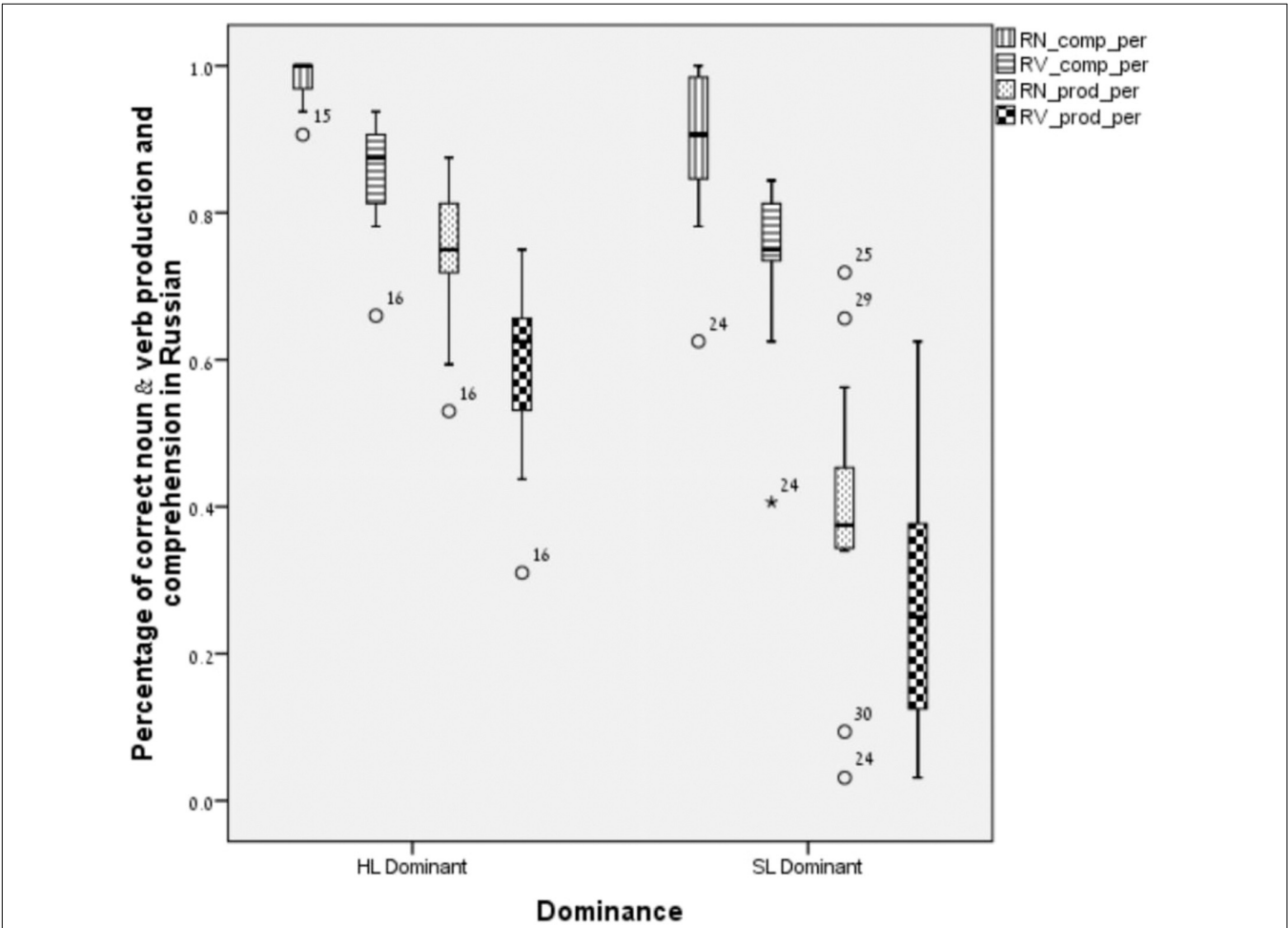
In order to explore whether vocabulary size is different in the two dominance groups, descriptive results on both receptive and expressive abilities of children on verbs and nouns in their HL (Russian) and SL (Hebrew) are presented. **Table 2** presents a comparison of the HL-dominant bilingual children to the SL-dominant bilingual children. Monolingual data is presented for SL only. **Figures 1, 2** present the group differences in HL and SL, respectively.

**Table 2** shows that vocabulary size mirrors the dominance level of the two groups. The children’s performance was better in the language in which they were dominant in terms of both receptive and expressive vocabulary.

For HL-Russian, a one-way MANOVA, with nouns and verbs receptive and expressive vocabulary scores in Russian as dependent variables, and language groups (SL-dominant vs. HL-dominant bilinguals) as an independent variable, was conducted. A significant multivariate effect was found for Language groups,  $F(4,31) = 17.47, p < 0.05$ ; Wilks’  $\lambda = 0.3, \eta^2 = 0.69$ , such that HL-dominant bilinguals outperformed

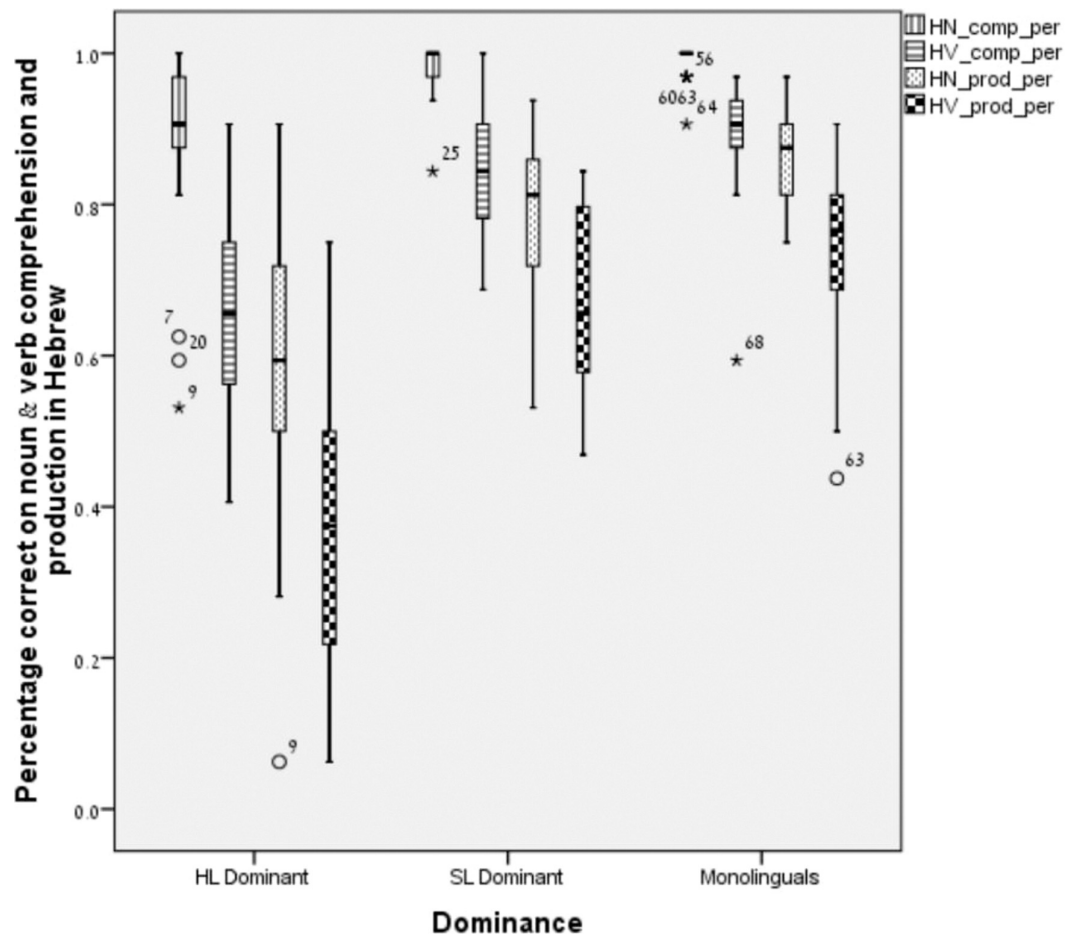
**TABLE 2 |** Vocabulary size via receptive and expressive vocabulary of nouns and verbs in the HL (Russian) and SL (Hebrew).

	HL-Russian vocabulary		SL-Hebrew vocabulary		Monolingual
	SL-dominant	HL-dominant	SL-dominant	HL-dominant	
Noun receptive	0.90 (0.12)	0.98 (0.03)	0.97 (0.04)	0.88 (0.14)	0.99 (0.02)
Verb receptive	0.76 (0.11)	0.85 (0.07)	0.85 (0.10)	0.66 (0.13)	0.89 (2.16)
Noun expressive	0.50 (0.24)	0.70 (0.16)	0.78 (0.12)	0.58 (0.21)	0.86 (0.06)
Verb expressive	0.36 (0.21)	0.55 (0.16)	0.68 (0.13)	0.35 (0.20)	0.73 (0.1)



**FIGURE 1 |** Noun and verb production and comprehension in HL-Russian.





**FIGURE 2 |** Noun and verb production and comprehension in SL-Hebrew.

SL-dominant bilinguals in receptive and expressive vocabulary in Russian (HL). Moreover, univariate testing indicated significant differences between the two language groups in each of the LITMUS CLT tasks: In the noun receptive task,  $F(1,34) = 12.76$ ,  $p < 0.01$ ,  $\eta^2 = 0.27$ ; in the verb receptive task,  $F(1,34) = 15.21$ ,  $p < 0.001$ ,  $\eta^2 = 0.31$ ; in noun expression,  $F(1,34) = 64.48$ ,  $p < 0.001$ ,  $\eta^2 = 0.65$ ; and in verb expression,  $F(1,34) = 42.48$ ,  $p < 0.001$ ,  $\eta^2 = 0.55$ . That is, there were significant differences between the two groups on all four vocabulary measures, with HL-dominant bilinguals outperforming SL-dominant bilinguals in receptive and expressive nouns and verbs in HL/Russian, as can be seen in **Figure 1**.

Likewise, for SL-Hebrew, an initial two-way MANOVA was conducted, with nouns and verbs receptive and expressive scores as dependent variables and language group (monolingual, SL-dominant, HL-dominant) as independent variables. Significant multivariate effect for language group,  $F(8,124) = 10.37$ ,  $p < 0.001$ ; Wilks'  $\lambda = 0.36$ ,  $\eta^2 = 0.4$ . A follow-up Bonferroni analysis showed that the average test score of monolinguals and dominant SL bilinguals was statistically higher than that of HL-dominant bilinguals in all four categories ( $p < 0.001$ ). There were no significant

differences between the monolinguals and the SL-dominant bilinguals. Moreover, univariate testing indicated significant differences between the two language groups in each of the LITMUS CLT tasks in Hebrew (SL): In the noun comprehension,  $F(1,34) = 7.11$ ,  $p < 0.05$ ,  $\eta^2 = 0.17$ ; in verbs comprehension,  $F(1,34) = 20.94$ ,  $p < 0.001$ ,  $\eta^2 = 0.38$ ; in nouns expression,  $F(1,34) = 11.53$ ,  $p < 0.01$ ,  $\eta^2 = 0.25$ ; and in verbs expression,  $F(1,34) = 32.52$ ,  $p < 0.001$ ,  $\eta^2 = 0.49$ . That is, there were significant differences between the two groups on all four vocabulary measures, with SL-dominant bilinguals outperforming HL-dominant bilinguals in receptive and expressive nouns and verbs in SL/Hebrew, as can be seen in **Figure 2**. Finally, there was a gradual pattern in all groups where the highest scores were found in noun receptive vocabulary followed by verb receptive vocabulary and only then did the expressive vocabulary follow with children performing higher on noun expressive vocabulary than on verb expressive vocabulary.

## Metalinguistic Awareness Measures

Metalinguistic awareness was measured in Hebrew. Descriptive results comparing the three groups' performances in the two

**TABLE 3 |** Morphological and lexical metalinguistic awareness tasks.

Metalinguistic awareness	HL-dominant	SL-dominant	Monolinguals
Morphological	0.62 (0.11)	0.71 (0.13)	0.72* (0.15)
Lexical	0.68 (29)	0.77 (0.29)	0.75 (0.21)

\* $p < 0.05$  for the difference between HL-dominant bilinguals and monolinguals.

metalinguistic awareness tasks (morphological and lexical) are presented in **Table 3**.

In order to examine whether there were differences between the three groups in the two metalinguistic awareness tasks, a one-way ANOVA was conducted. Significant differences were revealed in the morphological awareness task  $F(2,65) = 3.74$ ,  $p < 0.05$ . A *post hoc* Bonferroni analysis revealed that monolinguals outperformed HL-dominant bilinguals ( $p < 0.05$ ), with no significant differences between monolinguals and SL-dominant bilinguals or between HL-dominant and SL-dominant bilinguals. No Univariate effect was found for language groups in the lexical awareness task,  $F(2,64) = 0.70$ ,  $p > 0.05$ .

## Metalinguistic Awareness and Vocabulary Size

The major aim of the paper was to explore the relative contribution of dominance measured by relative proficiency, lexical and morphological metalinguistic awareness and the interaction between dominance and metalinguistic awareness in the SL-Hebrew to receptive and expressive vocabulary size in Hebrew in comparison to Russian. A hierarchical regression analysis was conducted with all five predictors, introducing relative proficiency first, followed by the metalinguistic awareness measures, and finally the interaction between relative proficiency and metalinguistic awareness.

## SL-Hebrew Receptive Vocabulary

**Table 4** presents a summary of the hierarchical regression analysis for variables predicting receptive lexicon in SL-Hebrew.

The hierarchical regression analysis shows that relative proficiency alone (Model 1) significantly predicted the size of

receptive vocabulary [ $\beta = 0.666$ ,  $t(34) = 5.134$ ,  $p < 0.001$ ]. Model 1 explained 42% of the variance in the size of receptive vocabulary [ $F(1,33) = 26.363$ ,  $p < 0.001$ ]. When metalinguistic awareness measures are added in Model 2, the model significantly predicted the size of receptive vocabulary [ $\beta = 0.604$ ,  $t(33) = 4.293$ ,  $p < 0.001$  for relative proficiency,  $\beta = 0.306$ ,  $t(33) = 2.458$ ,  $p = 0.02$  for lexical metalinguistic awareness], explaining together 49.7% of the variance [ $F(3,31) = 11.879$ ,  $p < 0.001$ ]. Morphological metalinguistic awareness made no significant contribution. When the interactions are added in Model 3, the new model significantly predicted the size of receptive vocabulary [ $\beta = 2.209$ ,  $t(32) = 3.084$ ,  $p = 0.004$  for relative proficiency,  $\beta = 0.376$ ,  $t(32) = 3.349$ ,  $p = 0.002$  for lexical metalinguistic awareness, and  $\beta = -1.290$ ,  $t(32) = -3.273$ ,  $p = 0.003$  for the interaction between relative proficiency and lexical metalinguistic awareness], explaining together 60.3% of the variance [ $F(5,29) = 11.348$ ,  $p < 0.001$ ]. Model 3 suggests that while relative proficiency and lexical metalinguistic awareness are positively related to the size of receptive vocabulary, the interaction between them is negatively related to the size of receptive vocabulary. Morphological metalinguistic awareness and the interaction between relative proficiency and morphological awareness have no significant contribution.

Due to the small number of bilingual participants, the hierarchical regression used above with five predictors is prone to overfitting. Therefore, we further conducted a linear regression for each dominance group in which only the two metalinguistic awareness measures were introduced as predictors. A similar linear regression was conducted for the monolingual group to provide a baseline for comparison. **Table 5** presents a summary of a linear regression analysis for the two variables predicting receptive vocabulary size for HL-dominant and SL-dominant bilinguals as well as monolinguals.

The linear regression showed that for HL-dominant bilinguals lexical metalinguistic awareness significantly predicted the size of receptive vocabulary [ $\beta = 0.658$ ,  $t(20) = 3.623$ ,  $p = 0.002$ ], while morphological metalinguistic awareness does not contribute. Lexical metalinguistic awareness explained 37.2% of the variance in the size of receptive vocabulary [ $F(2,18) = 6.928$ ,  $p = 0.006$ ]. For the SL-dominant group and the monolingual group, no predictors were found to contribute.

**TABLE 4 |** Summary of hierarchical regression analysis for variables predicting receptive vocabulary size ( $N = 35$ ).

Variable	Model 1			Model 2			Model 3		
	<i>B</i>	<i>SE B</i>	$\beta$	<i>B</i>	<i>SE B</i>	$\beta$	<i>B</i>	<i>SE B</i>	$\beta$
RelProf	0.024	0.005	0.666***	0.024	0.005	0.604***	0.081	0.026	2.209**
LexM				0.128	0.052	0.306*	0.158	0.047	0.376**
MorphM				0.044	0.136	0.045	0.099	0.125	0.101
RelProf $\times$ LexM							-0.059	0.018	-1.290**
RelProf $\times$ MorphM							-0.022	0.032	-0.409
$R^2$		0.444			0.535			0.662	
<i>F</i>		26.363***			11.879***			11.348***	

RelProf, Relative proficiency; LexM, Lexical metalinguistic awareness; MorphM, Morphological metalinguistic awareness. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . One child was excluded from the analysis since he was missing lexical awareness scores.

**TABLE 5 |** Summary of linear regression analyses for variables predicting receptive vocabulary size for HL-dominant, SL-dominant bilinguals, and monolinguals.

Variable	HL-Dominant [N = 21]			SL-Dominant [N = 14]			Monolinguals [N = 32]		
	B	SE B	$\beta$	B	SE B	$\beta$	B	SE B	$\beta$
LexM	0.270	0.074	0.658**	-0.061	0.052	-0.322	0.017	0.032	0.100
MorphM	0.314	0.192	0.296	0.112	0.117	0.267	0.063	0.045	0.252
R <sup>2</sup>		0.435			0.161			0.083	
F		6.928**			1.057			1.306	

LexM, Lexical metalinguistic awareness; MorphM, Morphological metalinguistic awareness. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . One child was excluded from the analysis since he was missing lexical awareness scores.

**TABLE 6 |** Summary of hierarchical regression analysis for variables predicting expressive vocabulary size (N = 35).

Variable	Model 1			Model 2			Model 3		
	B	SE B	$\beta$	B	SE B	$\beta$	B	SE B	$\beta$
RelProf	0.047	0.008	0.723***	0.040	0.008	0.617***	0.123	0.043	1.905**
LexM				0.257	0.081	0.347**	0.296	0.078	0.399**
MorphM				0.216	0.219	0.126	0.304	0.207	0.177
RelProf $\times$ LexM							-0.069	0.030	-855*
RelProf $\times$ MorphM							-0.047	0.052	-0.507
R <sup>2</sup>		0.522			0.644			0.704	
F		36.081***				18.717***		13.823***	

RelProf, Relative proficiency; LexM, Lexical metalinguistic awareness; MorphM, Morphological metalinguistic awareness. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

## SL-Hebrew Expressive Vocabulary

Similar results were observed for the expressive vocabulary. **Table 6** presents a summary of the hierarchical regression analysis for variables predicting the expressive vocabulary in SL-Hebrew.

The hierarchical regression analysis shows that relative proficiency alone (Model 1) significantly predicted the size of the expressive vocabulary [ $\beta = 0.723$ ,  $t(34) = 6.007$ ,  $p < 0.001$ ]. Model 1 explained 50.8% of the variance in the size of the expressive vocabulary [ $F(1,33) = 36.081$ ,  $p < 0.001$ ]. When metalinguistic awareness measures are added in Model 2, the model significantly predicted the size of the expressive vocabulary [ $\beta = 0.617$ ,  $t(33) = 5.015$ ,  $p < 0.001$  for relative proficiency,  $\beta = 0.347$ ,  $t(33) = 3.189$ ,  $p = 0.003$  for lexical metalinguistic awareness], explaining together 61% of the variance [ $F(3,31) = 18.717$ ,  $p < 0.001$ ]. Morphological metalinguistic awareness made no significant contribution. When the interactions are added in Model 3, the new model significantly predicted the size of the expressive vocabulary [ $\beta = 1.905$ ,  $t(32) = 2.846$ ,  $p = 0.008$  for relative proficiency,  $\beta = 0.399$ ,  $t(32) = 3.802$ ,  $p = 0.001$  for lexical metalinguistic awareness, and  $\beta = -0.855$ ,  $t(32) = -2.321$ ,  $p = 0.028$  for the interaction between relative proficiency and lexical metalinguistic awareness], explaining together 65.3% of the variance [ $F(5,29) = 13.823$ ,  $p < 0.001$ ]. Model 3 suggests that while relative proficiency and lexical metalinguistic awareness are positively related to the size of expressive vocabulary, the interaction between them is negatively related to the size of expressive vocabulary. Morphological metalinguistic awareness and the interaction between relative proficiency and morphological awareness have no significant contribution.

Due to the small number of bilingual participants, the regressions used above with five predictors is prone to overfitting. Therefore, we further conducted a simple linear regression for each dominance group in which only the two metalinguistic awareness measures were introduced as predictors. A similar linear regression was conducted for the monolingual group to provide a baseline for comparison. **Table 7** presents a summary of the simple regression analyses for the two variables predicting expressive vocabulary size for HL-dominant and SL-dominant bilinguals as well as monolinguals.

The linear regression showed that for HL-dominant bilinguals both lexical metalinguistic awareness and morphological metalinguistic awareness significantly predicted the size of expressive vocabulary [ $\beta = 0.596$ ,  $t(20) = 3.216$ ,  $p = 0.005$  and  $\beta = 0.401$ ,  $t(20) = 2.162$ ,  $p = 0.044$ , respectively]. The model explained 34.6% of the variance in the size of expressive vocabulary [ $F(2,18) = 6.285$ ,  $p = 0.009$ ]. For the SL-dominant group and the monolingual group, no predictors were found to contribute.

## Comparing SL-Hebrew Receptive and Expressive Vocabulary

The similarity in the impact of lexical metalinguistic awareness on receptive and expressive vocabulary size is further demonstrated in the scatter plots in **Figure 3**.

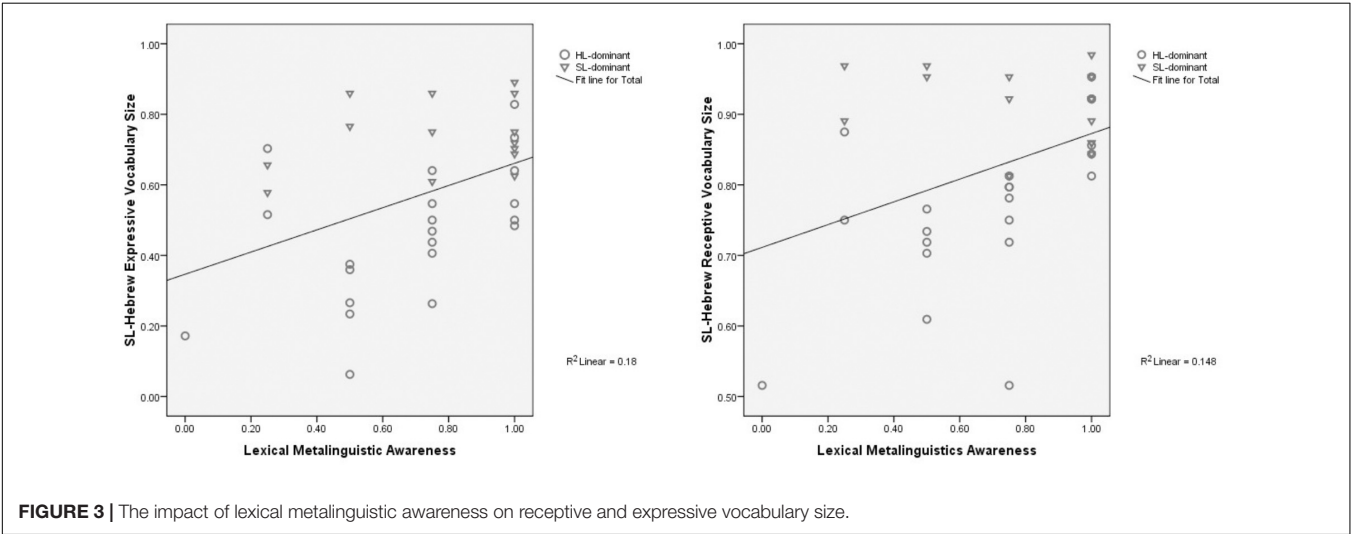
## HL-Russian Receptive and Expressive Vocabulary

The contribution of lexical metalinguistic awareness to vocabulary size in SL-Hebrew is in sharp contrast to the

**TABLE 7 |** Summary of linear regression analyses for variables predicting expressive vocabulary size for HL-dominant, SL-dominant bilinguals, and monolinguals.

Variable	HL-Dominant [N = 21]			SL-Dominant [N = 14]			Monolinguals [N = 32]		
	B	SE B	$\beta$	B	SE B	$\beta$	B	SE B	$\beta$
LexM	0.402	0.125	0.596**	0.087	0.099	0.245	−0.007	0.063	−0.021
MorphM	0.699	0.323	0.401*	0.233	0.220	0.293	0.055	0.091	0.114
R <sup>2</sup>		0.411			0.157			0.013	
F		6.285**			1.028			0.185	

LexM, Lexical metalinguistic awareness; MorphM, morphological metalinguistic awareness. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



findings for HL-Russian vocabulary size. Similar regression analyses conducted with HL-Russian receptive and expressive vocabulary size as the dependent variables, showed that only relative proficiency (which is positive for SL-dominant and negative for HL-dominant, by definition) negatively predicted vocabulary size in HL-Russian. Dominance measured by relative proficiency was the only predictor, explaining over 50% of the variance in receptive vocabulary, and over 70% of the variance in the expressive vocabulary. The metalinguistic awareness measures and the interactions were introduced in models 2 and 3, respectively, and had insignificant contribution.

DISCUSSION

The present study explored the possible connections between vocabulary size and different metalinguistic awareness abilities among bilingual children of different dominance groups and monolingual children with TLD. The first hypothesis that dominance, measured by relative proficiency, will impact vocabulary size in both languages was confirmed. Dominance groups differed in terms of vocabulary size. As expected, HL-dominant bilinguals outperformed SL-dominant bilinguals on SL-Russian receptive vocabulary. By contrast, SL-dominant bilinguals and monolinguals outperformed HL-dominant bilinguals on receptive and expressive vocabulary size in SL-Hebrew. For metalinguistic awareness, no difference was

found among the groups with one exception: monolinguals outperformed HL-dominant bilinguals on the morphological awareness tasks. When focusing on the different dominance groups, the linear regression showed that metalinguistic awareness abilities predicted vocabulary size only for the HL-dominant group, confirming the second hypothesis. Morphological metalinguistic awareness predicted vocabulary size only for expressive vocabulary among the HL-dominant group refuting the third hypothesis. The hierarchical regression analyses showed, that dominance, as well as lexical metalinguistic awareness and the interaction between the two, predicted receptive and expressive vocabulary size. Morphological metalinguistic awareness did not predict vocabulary size. This confirms the fourth hypothesis. Finally, no effect of metalinguistic awareness on HL, Russian vocabulary size, was observed for either group.

Receptive and Expressive Vocabulary

The results of the LITMUS CLT vocabulary task are in line with previous findings (e.g., Bialystok et al., 2010), with dominant HL bilinguals lagging behind their age-matched dominant SL and monolingual peers on all four vocabulary measures in SL (Hebrew), but outperforming their SL-dominant peers on all four vocabulary measures in HL (Russian). The lack of differences between monolinguals and SL-dominant bilinguals in vocabulary size is not surprising, considering the relative exposure to SL of the SL-dominant group



( $M = 54$  months), most of whom are simultaneous bilinguals. The regression analysis further showed that dominance measured by relative proficiency was the best predictor of receptive vocabulary among bilingual children confirming the first hypothesis.

Moreover, all children in this study had more difficulty with the expressive tasks than with the receptive tasks, as is consistent with previous literature (e.g., Gibson et al., 2012), with a bigger gap between receptive and expressive vocabulary size in SL-Hebrew for the bilingual groups (especially amongst HL-dominant) as opposed to the monolingual group. The expected receptive-expressive gap reflects the difference between the two processes. Receptive vocabulary that taps into lexical knowledge is less sensitive to language dominance than expressive vocabulary, that taps on lexical knowledge and its retrieval. This reflects the impact of dominance on more demanding processes (expression). The gap is consistently smaller in the dominant language (Russian in the HL-dominant bilinguals and SL-Hebrew in the SL-dominant bilinguals) than in the weaker language. Moreover, SL-dominant bilinguals perform like monolinguals. The sensitivity of lexical access to dominance suggests that the competition between the two linguistic representations of each concept is influenced by the relative proficiency in each language. While the receptive vocabulary is similar, the smaller gap in the dominant language suggests that linguistic representation in this language is more readily available in lexical access.

## Metalinguistic-Awareness Abilities

Children demonstrate metalinguistic awareness in later stages of language development, around the age of 5–6, after gradually mastering the structure of the language, accumulating vocabulary, and developing efficient access to words and concepts (Duncan et al., 2009). The present study shows no differences between the three groups of 6-year-olds in terms of metalinguistic awareness, except for one instance where monolinguals did significantly better than HL-dominant bilinguals on a morphological awareness task. Russian and Hebrew have very distinct morphological features, especially in word formation. Russian word formation highly relies on concatenative morphology (Shevelov, 1957), while Hebrew word formation mostly uses non-concatenative morphology (Berman and Bolozky, 1978; Aronoff, 1994). Previous studies suggested that morphological awareness requires high proficiency in a given language (Bialystok and Barac, 2012); thus, morphological awareness in SL-Hebrew requires high proficiency in SL-Hebrew. The finding that monolinguals outperformed the HL-dominant bilinguals on the morphological awareness task is in line with this assumption.

As the morphological task in this study depended on knowledge of SL-Hebrew derivational morphology, and knowledge of Hebrew derivational morphology requires, in turn, extensive knowledge of vocabulary, the limited Hebrew vocabulary size of HL-dominant children can be responsible for the gap in morphological metalinguistic awareness. A possible support for this explanation comes from the performance of the SL-dominant bilinguals. The SL-dominant bilinguals often

patterned with the monolinguals showing a significant difference from HL-dominant bilinguals. By contrast, for morphological awareness, they showed no significant differences from monolinguals as well as HL-dominant bilinguals, performing in between the two groups.

An explanation of the relatively limited morphological awareness abilities of HL-bilinguals could be their relatively low length of exposure to the SL, a variable that has great impact on language proficiency for bilingual children (e.g., Chondrogianni and Marinis, 2011). It is possible that HL-dominant bilinguals, who are often sequential bilinguals, did not have sufficient exposure (12–34 months) to their SL (Hebrew) in order to develop high morphological awareness in this language. Yet, the absence of a significant difference from the SL-dominant group that has longer exposure undermines this explanation. A definite conclusion on this is hampered by the small sample of children in the SL-dominant bilingual group ( $N = 15$ ) and the considerable variance in the length of exposure of the group ( $M = 54$ ,  $SD = 28.66$ ), which might have resulted in the lack of statistical differences between SL-dominant bilinguals and the other two groups.

Finally, the lack of difference between the groups in lexical awareness might have to do with the task selected for the present study. Lexical awareness was assessed through a fast mapping task. Fast mapping requires children of the age tested to consult their vocabulary when encountering a new word in order to meet the requirement of assigning a novel label to a novel object on the one hand and abide by conventionality on the other. Fast mapping resembles the situation often encountered in language learning by monolinguals (mapping a novel word form to a novel object). In bilingual language learning, the novel word in the SL is mapped onto a known object with a known label in the HL and does not follow mutual exclusivity. The lack of difference between the groups suggests that bilingual experience does not impact fast mapping as a measure of lexical metalinguistic awareness.

## The Relation Between Vocabulary Size and Metalinguistic Awareness

Better metalinguistic skills are expected to positively impact the acquisition of the SL. Hierarchical regression analyses tested the impact of dominance, measured by relative proficiency, lexical metalinguistic awareness, morphological metalinguistic awareness and the interaction of the two with dominance on the size of receptive and expressive vocabulary in both the HL and the SL. These analyses showed that beyond the significant impact of dominance, lexical metalinguistic awareness, but not morphological awareness, influenced vocabulary size. Despite the gap between receptive and expressive vocabulary, the impact of metalinguistic awareness was similar in the two modalities. The contribution of lexical metalinguistic awareness to vocabulary size among bilingual children suggests that bilinguals, like monolinguals, rely on fast mapping in expanding their vocabulary size. More specifically, the principles that are operative beyond early childhood for consciously monitoring the learning of novel words (Ramachandra et al., 2010) were found to be related to expanding the lexicon in the SL. The awareness of constraints on mapping novel names to nameless

objects to meet mutual exclusivity, seems to help in mapping novel words in the SL to objects, even if they already have a name in the HL. Likewise, the consideration of the use of conventional names for referents, seems to not block the process of mapping a novel name in one language to familiar objects that already have a conventional name in the other. This suggests that the utilization of the principles of fast mapping is sensitive to the language that is acquired. Having a label for an object in one language does not interfere with acquiring a new label in the other.

Our findings even suggest that experience with fast mapping, which is language neutral, helps in increasing the size of the lexicon. Morphological awareness, by contrast, was found to make little contribution, especially when the interaction between dominance and metalinguistic awareness was considered in the equation. These findings suggest that the language specific nature of morphological awareness tasks makes it impossible to rely on experience in one language in learning new words in the other.

Moreover, the significant contribution of lexical metalinguistic awareness to vocabulary size was limited to the SL-Hebrew, and was not observed in the HL-Russian. Experience with fast mapping, which is language neutral, seems to be transferred from the HL to the SL and helps in increasing the vocabulary size in the SL only. This asymmetry reflects the different phase each group is in for vocabulary acquisition in the two languages. A large number of the bilinguals in this study had a smaller vocabulary size in SL-Hebrew compared to HL-Russian. This suggests that they need to learn new vocabulary items at a more rapid speed in SL-Hebrew than in the HL-Russian. In such a case, better fast mapping skills can become useful.

This latter proposal is supported by the findings of the linear regression that lexical awareness was found to influence SL vocabulary size only in the HL-dominant group. While higher relative proficiency and greater lexical metalinguistic awareness was related to greater receptive and expressive vocabulary in SL-Hebrew, there was also an interaction between proficiency and lexical awareness. This interaction showed that the relationship between lexical awareness and vocabulary size was stronger for participants with lower proficiency. This supports the assumption that better lexical awareness, and in particular better fast mapping skills, predicts growth in vocabulary size, in different ways for different relative proficiency levels. In particular, this confirmed our second hypothesis that fast mapping which is important to lexical growth will show a stronger relation to vocabulary size at earlier stages in acquisition, that is, in the less dominant language.

These results were further confirmed by the linear regression conducted when focusing on each dominance group separately. For the different dominance groups, the regression analyses revealed that children rely on this metalinguistic ability if the SL is their less dominant language. The task used for lexical awareness predicts success in acquiring a larger vocabulary among the least proficient group, strengthening the above explanation, and showing the importance of introducing relative proficiency into the equation. The relationship between Hebrew vocabulary size and lexical awareness ability was found only

among HL-dominant bilinguals, but not for the other groups. The absence of such a relationship among the SL-dominant bilinguals is reminiscent of Kan and Kohnert's (2008) findings. There, they tested the relationship between lexical awareness (via fast mapping) and vocabulary size in both the HL and the SL (English) of sequential bilingual children with TLD, aged 3–5 and found that there were no significant correlations between vocabulary size and fast mapping across the two languages. Our SL-dominant bilingual children seem to be at the same stage of vocabulary acquisition as the children in Kan and Kohnert's (2008) study were. As the HL-dominant bilinguals are at the earlier stage of vocabulary acquisition, they still rely on these abilities, while the SL-dominant bilinguals and monolingual children are beyond this phase and therefore present a different profile. In sum, our findings suggest that metalinguistic awareness might have a different effect on vocabulary size at different levels of acquisition, which is consistent with the previous literature that shows different cognitive mechanisms operating at different stages of language acquisition (Gathercole et al., 1992; Hu, 2008).

Our findings for morphological metalinguistic awareness can also shed light on the question of whether metalinguistic awareness depends on the stage of language acquisition of SL that each group is at. Metalinguistic awareness might be limited by restricted formal linguistic knowledge in a particular language (Bialystok et al., 2014) and the stage in which each group is at in their language acquisition of SL. There are reasons to assume that the outcomes of this study, and in particular the negative relation observed among HL-dominant children between morphological awareness and their HL-vocabulary size, are related to their limited exposure to Hebrew morphology used in the relevant metalinguistic tasks. A task that will add measures of metalinguistic abilities in the HL will enable more definite conclusions.

To conclude, this study highlights the importance of considering dominance when studying language abilities and metalinguistic awareness among bilinguals. This is important in order to provide a more accurate account of the impact of bilingualism and better our understanding of the contribution of the relative proficiency in each language in each modality (expressive and receptive) and of metalinguistic awareness to vocabulary growth among bilinguals. A strong similarity was found between SL-dominant and monolingual children in SL vocabulary size while HL-dominant bilinguals lagged behind. By contrast, HL-dominant bilinguals outperformed SL-dominant bilingual on HL vocabulary size. The novelty of this study lies in the finding that the relation between metalinguistic awareness and vocabulary size were different in the two dominance groups. The HL-dominant group presented an earlier phase in the acquisition of the SL, in which vocabulary size in the SL is sensitive to lexical awareness, while vocabulary size in the HL hinders the development of morphological awareness in the SL. HL-dominant bilinguals relied on lexical metalinguistic awareness, measured by fast mapping abilities in expanding their vocabulary size, whereas SL-dominant, like monolinguals, did not. This shows that lexical awareness is important for word learning at more initial stages of vocabulary

acquisition. While many studies show the relevance of length and amount of exposure to vocabulary size, the present study shows that metalinguistic awareness should also be taken into consideration, and might make different contributions in different dominance groups.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of ethics guidelines of Bar Ilan University's IRB. The protocol was approved by the Bar Ilan University IRB as well as by the ethics committee at the Ministry of Education in Israel. All parent gave written informed consent and children gave their assent orally in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

CA, TG, and SA-L were responsible for the conception, analysis, and interpretation of the work. The paper was drafted and revised to include intellectual content by CA, TG, and SA-L. CA is accountable for the integrity and accuracy of the

work. The paper was approved for publication of content by SA-L.

## FUNDING

This research was supported by The Israel Science Foundation (Grant No. 863/14) and by a grant from the Ministry of Education.

## ACKNOWLEDGMENTS

This research was initiated within COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistics Patterns and the Road to Assessment" ([www.bi-sli.org](http://www.bi-sli.org)) and used the design developed by Working Group 3 under the leadership of Ewa Haman and Shula Chait. We thank Efrat Harel for her major contribution to the development of LITMUS-CLT-Hebrew. We thank the research assistants at Bar Ilan University, and in particular, Sveta Fichman and Sharon Granner. We would also like to thank the two reviewers for their contribution to the final outcome of the manuscript.

## REFERENCES

- Adesope, O., Lavin, T., Thompson, T., and Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review. of Educational. Research.*, 80, 207–245. doi: 10.3102/0034654310368803
- Altman, C., Armon-Lotem, S., Fichman, S., and Walters, J. (2016). Macrostructure, microstructure and mental state terms in the narratives of English-Hebrewbilingual preschool children with and without SLI. *Applied. Psycholinguistics.* 37, 165–193. doi: 10.1017/S0142716415000466
- Altman, C., Goldstein, T., and Armon-Lotem, S. (2017). Quantitative and qualitative differences in the lexical knowledge of monolingual and bilingual children. *Clinical. Linguistics. and Phonetics.*, 31 (11–12), 931–954. doi: 10.1080/02699206.2017.1312533
- Armon-Lotem, S. (2014). Between L2 and SLI: inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. *J. Child Lang.* 41, 3–33. doi: 10.1017/S0305000912000487
- Armon-Lotem, S., and Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of Specific Language Impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal. of Language. and Communication. Disorders.*, 51, 715–731 doi: 10.1111/1460-6984.12242
- Armon-Lotem, S. de Jong, J., and Meir, N. (eds). (2015). *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. Bristol: Multilingual matters.
- Aronoff, M. (1994). *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, MA: MIT press.
- Aşık, A., Vural, A. S., and Akpınar, K. D. (2015). Lexical awareness and development through data driven learning: attitudes and beliefs of EFL learners. *J. Journal of Education. and Training. Studies.*, 4(3), 87–96. doi: 10.11114/jets.v4i3.1223
- Behrend, D. A., Scofield, J., and Kleinknecht, E. E. (2001). Beyond fast mapping: young children's extensions of novel words and novel facts. *Dev. Psychol.* 37(5), 698–705. doi: 10.1037//0012-1649.37.5.698
- Berman, R. A., and Bolozky, S. (1978). *Modern Hebrew Structure*. Tel Aviv: University Pub Projects.
- Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511605963
- Bialystok, E., and Barac, R. (2012). Emerging bilingualism: dissociating advantages for metalinguistic awareness and executive control. *Cognition* 122, 67–73. doi: 10.1016/j.cognition.2011.08.003
- Bialystok, E., Luk, G., and Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: interactions among languages and writing systems. *Scientific. Studies. of Reading.* 9, 43–61. doi: 10.1207/s1532799xssr0901\_4
- Bialystok, E., Luk, G., Peets, K. F., and Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13, 525–553. doi: 10.1017/S1366728909990423
- Bialystok, E., McBride-Chang, C., and Luk, G. (2005b). Bilingualism, language proficiency and learning to read in two writing systems. *J. Educ. Psychol.* 97, 580–590. doi: 10.1037/0022-0663.97.4.580
- Bialystok, E., Peets, K.F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied. Psycholinguistics.* 35, 177–191.
- Bowey, J. A. (1986). Syntactic awareness and verbal performance from pre-school to fifth grade. *J. Psycholinguist. Res.* 15, 285–308. doi: 10.1007/BF01067676
- Braisby, N., Dockrell, J. E., and Best, R. M. (2001). "Children's acquisition of science terms: does fast mapping work"? in *Research on Child Language Acquisition: Proceedings of the 8th Conference of the International Association for the Study of Child Language*. eds M. Almgren, A. Barreña, M. J. Ezeizabarrena, I. Idiazabal, and B. MacWhinney (Somerville, MA: Cascadia Press), 1066–1087. Available at: <http://discovery.ucl.ac.uk/id/eprint/10026727>
- Carey, S., and Bartlett, E. (1978). Acquiring a single new word. *Papers. and Reports. on Child Language. Development.* 15, 17–29.
- Carlisle, J. F., Beeman, M., Davis, L. H., and Spharim, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied. Psycholinguistics.* 20, 459–478. doi: 10.1017/S0142716499004014
- Chen, X., Hao, M., Geva, E., Zhu, J., and Shu, H. (2009). The role of compound awareness Chinese children's vocabulary acquisition and character reading. *Reading. and Writing.* 22, 615–631. doi: 10.1007/s11145-008-9127-9
- Chen, X., Ramirez, G., Luo, Y. C., Geva, E., and Ku, Y. (2012). Comparing vocabulary development in Spanish and Chinese-speaking ELLs: the effects of metalinguistic and sociocultural factors. *Reading. and Writing.* 25, 1991–2020. doi: 10.1007/s11145-011-9318-7



- Cheung, H., Chung, K. K. H., Wong, S. W. L., McBride-Chang, C., Penney, T. B. and Ho, C. S. (2010). Speech perception, metalinguistic awareness, reading, and vocabulary in Chinese-English bilingual children. *Journal of Educational Psychology*, 102, 367–380. doi: 10.1037/a0017850
- Chondrogiani, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, 1, 318–345. doi: 10.1075/lab.1.3.05cho
- Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: analysis and control in young Swedish–English bilinguals. *Applied Psycholinguistics*, 20, 1–20. doi: 10.1017/S0142716499001010
- Duncan, L. G., Casalis, S., and Cole, P. (2009). Early metalinguistic awareness of derivational morphology: observations from a comparison of English and French. *Applied Psycholinguistics*, 30(3), 405–440. doi: 10.1017/S0142716409000213
- Farnia, F., and Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics*, 32(4), 711–738
- Foroodi Nejad, F., and Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Farsi-English bilinguals. *Bilingualism: Language and Cognition*, 12, 411–427. doi: 10.1017/S1366728909990241
- Fromkin, V., Rodman, R., and Hyams, N. M. (2007). *An Introduction to Language*, 8rd 8th Edn. Nashville, TN: Australia: Thomson Wadsworth.
- Gagarina, N., Klassert, A., and Topaj, N. (2010). *Russian Language Proficiency Test for Multilingual Children*. ZAS Papers in Linguistics 54, Berlin: ZAS.
- Gagarina, N., and Nenonen, O. (2017). Cross-linguistic Lexical Tasks: Russian version (CLT- RU). Part of COST IS0804 LITMUS Battery\*. Unpublished material. 3:43.
- 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. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *BilingualismLang. Cogn.* 12, 213–237. doi: 10.1017/S1366728909000415
- Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012) The receptive–expressive gap in the vocabulary of young second-language learners: robustness and possible mechanisms. *Bilingualism: Language and Cognition*, 15(1), 102–106. doi: 10.1017/S1366728910000490
- Goldstein, B. A., Fabiano, L., and Washington, P. S. (2005). Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools*, 36, 201–218.
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., and Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Dev. Psychol.* 28, 99–108. doi: 10.1037/0012-1649.28.1.99
- Golinkoff, R. M., Mervis, C. B., and Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *J. Child Lang.* 21, 125–155.
- Goralnik, E. (1995). *Goralnik Screening Test for Hebrew*, . Even Yehuda: Matan.
- Gray, S. (2006). The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *J. Speech Lang. Hear. Res.* 49, 955–969.
- Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.
- Gross, M., Buac, M., and Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *Am. J. Speech Lang. Pathol.* 23, 574–586. doi: 10.1044/2014\_AJSLP-13-0026
- Guo, Y., Roehrig, A. D., and Willia, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, 43(2), 159–183. doi: 10.1177/1086296X11403086
- Haman, E., Łuniewska, M., and Pomiechowska, B. (2015). “Designing cross-tasks (CLTs) for bilingual preschool children”. in *Methods for Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: disentangling bilingualism from Language impairment*, (Bristol: Multilingual Matters.), 194–238.
- Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., and Goldstein, B. (2012). Predicting Spanish–English bilingual children's language abilities. *J. Speech Lang. Hear. Res.* 55, 1251–1264. doi: 10.1044/1092-4388(2012)11-0016
- 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), 1–27. doi: 10.1017/S0305000910000759
- Hu, C. -F. (2003). Phonological memory, phonological awareness, and Foreign Language word learning. *Language Learning*, 53, 429–462. doi: 10.1111/1467-9922.00231
- Hu, C. -F. (2008). Rate of acquiring and processing L2 color words in relation to L1 phonological awareness. *The Modern Language Journal*, 92, 39–52. doi: 10.1111/j.1540-4781.2008.00685.x
- Hu, C. -F., and Schuele, C. M. (2005). Learning nonnative names: the effect of poor native phonological awareness. *Applied Psycholinguistics*, 26, 343–362. doi: 10.1017/S0142716405050204
- Ibrahim, R., Eviatar, Z., and Aharon-Peretz, J. (2007). Metalinguistic awareness and reading performance: a cross-language comparison. *J. Psycholinguist. Res.* 36, 297–317. doi: 10.1007/s10936-006-9046-3
- Iluz-Cohen, P., and Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism: Language and Cognition*, 16, 884–899. doi: 10.1017/S1366728912000788
- Kan, P. F., and Kohnert, K. (2005). Preschoolers learning Hmong and English: lexical-semantic skills in L1 and L2. *J. Speech Lang. Hear. Res.* 48, 372–383. doi: 10.1044/1092-4388(2005)026
- Kan, P. F., and Kohnert, K. (2008). Fast mapping by bilingual preschool children. *J. Child Lang.* 35, 495–514. doi: 10.1017/S0305000907008604
- Kan, P. F., Sadagopan, N., Janich, L., and Andrade, M. (2014). Effects of speech practice on fast mapping in monolingual and bilingual speakers. *J. Speech Lang. Hear. Res.* 57, 929–941. doi: 10.1044/2013\_JSLHR-L-13-0045
- Keller, K., Troesch, L. M., and Grob, A. (2015). A large receptive-expressive gap in bilingual children. *Frontiers in Psychology*, 6, :1284. doi: 10.3389/fpsyg.2015.01284
- Kraut, R. (2015). The relationship between morphological awareness and morphological decomposition among English language learners. *Reading and Writing*, 28(6), 873–890. doi: 10.1007/s11145-015-9553-4
- Kuo, L. J., and Anderson, R. C. (2006). Morphological awareness and learning to read: a cross-language perspective. *Educational Psychologist*, 41, 161–180. doi: 10.1207/s15326985sep4103\_3
- Marecka, M., Szewczyk, J., Jelec, A., Janiszewska, D., Rataj, K., and Dziubalska-Kolaczyk, K. (2018). Different phonological mechanisms facilitate vocabulary learning at early and late stages of language acquisition: evidence from Polish 9-year-olds learning English. *Applied Psycholinguistics*, 39, 1–35. doi: 10.1017/S0142716417000455
- McBride-Chang, C., Wagner, R. K., Muse, A., Chow, B. W.-Y., and Shu, H. (2005). The role of morphological awareness in children's vocabulary acquisition in English. *Applied Psycholinguistics*, 26, 415–435. doi: 10.1017/S014271640505023X
- Meir, N., Walters, J., and Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: the impact of L1 and L2 properties. *International Journal of Bilingualism*, 20, 421–452. doi: 10.1177/1367006915609240
- Mervis, C. B., and Bertrand, J. (1994). Acquisition of the novel name–nameless category (N3C) principle. *Child Dev.* 65, 1646–1662. doi: 10.2307/1131285
- Montrul, S. (2016). “Dominance and proficiency in early and late bilingualism,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35. doi: 10.1017/CBO9781107375345.002
- Nagy, W., Berninger, V. W. and Abbott, R., Vaughan, K., and Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology*, 95, 730–742. doi: 10.1037/0022-0663.95.4.730
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge, England: Cambridge University Press. doi: 10.1017/CBO9781139524759
- Nation, P. (2008). “Lexical awareness in second language learning,” in *Encyclopedia of Language and Education*, ed. N. H. Hornberger (Boston, MA: Springer), 1924–1934. doi: 10.1007/978-0-387-30424-3\_147



- O'Toole, C., Gatt, D., Hickey, T. M., Miéksz, A., Haman, E., Armon-Lotem, S., and Kern, S. (2017). Parent report of early lexical production in bilingual children: a cross-linguistic CDI comparison. *International Journal of Bilingual Education and Bilingualism*, 20, 124–145. doi: 10.1080/13670050.2016.1179258
- Páez, M., and Rinaldi, C. (2006). Predicting English word reading skills for Spanish-speaking students in first grade. *Topics in Language Disorders*, 26:338. doi: 10.1097/00011363-200610000-00006
- Páez, M. M., Tabors, P. O., and López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *Journal of Applied Developmental Psychology*, 28, 85–102. doi: 10.1016/j.appdev.2006.12.007
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., and Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18, 41–58. doi: 10.1017/S0142716400009863
- Ramachandra, V., Rickenbach, B., Ruda, M., LeCureux, B., and Pope, M. (2010). Fast mapping in healthy young adults: the influence of metamemory. *J. Psycholinguist. Res.* 39, 213–224. doi: 10.1007/s10936-009-9133-3
- Ramirez, G., Walton, P., and Roberts, W. (2013). Morphological awareness and vocabulary development among kindergarteners with different ability levels. *J. Learn. Disabil.* 47(1), 54–64. doi: 10.1177/0022219413509970
- Raven, J. C. (1938). *Raven's Progressive Matrices*. Los Angeles, CA: Western Psychological Services.
- Reder, F., Marec-Breton, N., Gombert, J., and Demont, E. (2013). Second-language learners' advantage in metalinguistic awareness: a question of languages' characteristics. *British Journal of Educational Psychology*, 83, 686–702. doi: 10.1111/bjep.12003
- Shevelov, G. Y. (1957). The structure of the root in modern Russian. *The Slavic and East European Journal*, 1, 106–124. doi: 10.2307/304136
- Spaulding, T. J., Hosmer, S., and Schechtman, C. (2013). Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *International Journal of Speech-Language Pathology*, 15, 453–463. doi: 10.3109/17549507.2012.762042
- Tabors, P. O., Paez, M., and López, L. M. (2003). Dual language abilities of bilingual four-year olds: initial findings from the early childhood study of language and literacy development of spanish-speaking children. *NABE Journal of Research and Practice*, 1, 170–1091.
- Uccelli, P., and Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, 38, 225–236.
- Umbel, V. M., and Ki Oller, D. (1994). Developmental changes in receptive vocabulary in Hispanic bilingual school children. *Language Learning*, 44, 221–242. doi: 10.1111/j.1467-1770.1994.tb01101.x
- Unsworth, S. (2015). "Amount of exposure as a proxy for dominance in bilingual language acquisition." in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds J. Treffers-Daller, J., and C. Silva-Corvalán, C. (Eds.). Language dominance in bilinguals: Issues of measurement and operationalization. (Cambridge: Cambridge University Press).
- Verhoeven, L., van Leeuwe, J., and Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, 15, 8–25. doi: 10.1080/10888438.2011.536125
- Winsler, A., Díaz, R. M., Espinosa, L., and Rodríguez, 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. (2007). *The Bilingual Child: Early Development and Language Contact*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620744

**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 Altman, Goldstein and Armon-Lotem. 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.



# Different Outcomes in the Acquisition of Residual V2 and *Do*-Support in Three Norwegian-English Bilinguals: Cross-Linguistic Influence, Dominance and Structural Ambiguity

Merete Anderssen\* and Kristine Bentzen\*

Department of Language and Culture, UiT – The Arctic University of Norway, Tromsø, Norway

## OPEN ACCESS

### Edited by:

Esther Rinke,  
Goethe-Universität Frankfurt am Main,  
Germany

### Reviewed by:

Cecilia Poletto,  
Goethe-Universität Frankfurt am Main,  
Germany

Michael Zimmermann,  
Universität Konstanz, Germany

### \*Correspondence:

Merete Anderssen  
merete.anderssen@uit.no  
Kristine Bentzen  
kristine.bentzen@uit.no

### Specialty section:

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

**Received:** 25 June 2018

**Accepted:** 16 October 2018

**Published:** 09 November 2018

### Citation:

Anderssen M and Bentzen K  
(2018) Different Outcomes  
in the Acquisition of Residual V2  
and *Do*-Support in Three  
Norwegian-English Bilinguals:  
Cross-Linguistic Influence,  
Dominance and Structural Ambiguity.  
Front. Psychol. 9:2130.  
doi: 10.3389/fpsyg.2018.02130

This paper investigates the acquisition of residual verb second (V2) in three corpora consisting of data from Norwegian-English bilinguals (Emma, Emily and Sunniva) in order to determine to what extent these structures are affected by cross-linguistic influence (CLI) from Norwegian V2. The three girls exhibit three different patterns with regard to the relevant constructions. They are very target-like in their use of auxiliaries in the relevant structures. However, when it comes to *do*-support, Emily and Sunniva are equally target-like, while Emma mainly produces non-target-like structures. These either involve the omission of *do*, or non-target-like movement of a lexical verb. Furthermore, Emma also allows verb movement across the subject with both lexical verbs and auxiliaries in topicalised structures, suggesting that she has overgeneralised residual V2 across verb types and clause types. Emily, on the other hand, is very target-like in structures involving residual V2 in English, but also allows auxiliaries and dummy-*do* to move across the subject in topicalised structures, overgeneralising residual V2 to apply to non-subject-initial declaratives. Finally, Sunniva is very precocious and very target-like in all the relevant structures, which may be an indication of acceleration due to CLI from Norwegian V2. We discuss these results with reference to language balance, finding that the measures available to us suggest that the differences between the children cannot straightforwardly be explained by language dominance. Instead, we suggest that these results can be accounted for by ambiguity in the English system, leaving the data open to several possible interpretations when acquired in contact with the consistent V2 system in Norwegian. This has several consequences: (i) the three girls' parsers interpret the input differently, (ii) differences between the three children are qualitative rather than quantitative and (iii) there has to be some mechanism that ensures that the children can 'recover' from these non-target-like grammars. In this paper, we will focus on the first two issues.

**Keywords:** bilingualism, English, Norwegian, *do*-support, verb second, residual verb second, cross-linguistic influence, language dominance

## INTRODUCTION

While it is generally agreed that bilingual children separate their two languages from very early on (cf. De Houwer, 2009 for an overview), it is also clear that the two languages of bilingual children may influence each other. Cross-linguistic influence (CLI) is indeed a typical characteristic of bilingual first language acquisition (cf. Serratrice, 2013 for an overview). CLI may have several potential consequences, the most common one being a delay in the acquisition of a particular feature (e.g. Sorace, 2005; Patuto et al., 2011). However, CLI has also been shown to result in a developmental path for bilinguals that *diverges* from that found in monolinguals (e.g. Anderssen and Bentzen, 2013) and in some cases also leading to accelerated development, sometimes in combination with language dominance (e.g. in Kupisch and Bernardini, 2007) but not always (Liceras et al., 2011).

Concerning the underlying causes for CLI, various sources have been explored. Hulk and Müller (2000) and Müller and Hulk (2001) proposed that the syntax-pragmatics interface was particularly vulnerable to CLI. This proposal has been further developed in work by among others Serratrice et al. (2004, 2009), Sorace et al. (2009) and Liceras et al. (2011). Moreover, Hulk and Müller also proposed that CLI would be more plausible in contexts where the two languages display superficial structural overlap. This may lead the child to pose parallel structural analyses to a certain construction in the two languages even in cases where the two languages actually are underlyingly different. Finally, language dominance is another factor that has been explored as a cause for CLI, in particular in explaining the *direction* of CLI (e.g. Genesee et al., 1995; Bernardini and Schlyter, 2004; Foroodi-Nejad and Paradis, 2009).

In the current paper, we address the cause and nature of CLI by investigating the acquisition of *do*-support and residual V2 in three English-Norwegian bilingual children. There is substantial superficial structural overlap with respect to word order between English and Norwegian, as both languages have a basic SVO word order. However, Norwegian is a V2 language, and as a result, *all* finite verbs consistently move to the second position in matrix clauses. English, on the other hand, is not a V2 language but nevertheless has a number of structures exhibiting V2-like characteristics, and it is often referred to as residual V2 (Rizzi, 1996).

The relevant contexts we investigate in this paper are illustrated in (1)–(5). English has V-to-T movement with auxiliaries and *be*, which is visible in clauses with negation or adverbials (1). Moreover, in interrogative clauses, English displays Subject Auxiliary Inversion (SAI) (2). However, in both negative and interrogative clauses, *do*-support is required in the absence of a finite auxiliary [(3), (4)]. In non-subject-initial declaratives, so-called topicalised constructions, English does not display SAI (5).

- (1) I *have* not seen her.
- (2) *Have* you seen her?
- (3) I *did* not buy the book.

- (4) *Did* you buy the book?
- (5) (a) Today I *have* bought a new car.  
(b) Yesterday I watched the new Star Wars movie.

The three bilingual girls in this study display three different patterns with respect to the acquisition of residual V2, *do*-support and non-subject-initial clauses in English that all diverge from what is typically found in monolingual English-speaking children. We will therefore explore whether all these three outcomes of the bilingual situation are due to CLI from Norwegian V2.

The paper is structured as follows. In the ‘Background’ section, we first provide the relevant background on target-like *do*-support and verb placement in English and Norwegian. We highlight where the two languages display superficially overlapping surface structures, and where the two systems are underlyingly (and superficially) different, thus pointing out where CLI due to structural overlap might be expected. In the ‘Previous Research on the Acquisition of Auxiliaries and *Do*-Support in English’ section, we present previous research on *do*-support and verb placement in monolingual English-speaking children, as well as previous research on the acquisition of verb placement in children acquiring English alongside a (Germanic) V2 language. Finally, in the ‘Research Questions and Predictions for the Current Study’ section, we present the research questions of the current paper. In the ‘Materials and Methods’ section, we introduce the three bilingual girls studied in this investigation, as well as our methodology. In the ‘Results’ section, we present the results of our investigation. In the ‘Discussion’ section contains a discussion of the results, and we explore to what extent the differences between the three girls can be attributed to language dominance. In the ‘Conclusion’ section concludes the paper.

## BACKGROUND

### Verb Placement in Norwegian and English

In this section, we outline the crucial background on verb placement in Norwegian and English, highlighting areas of superficial structural overlap that might be susceptible to CLI.

#### Verb Second in Norwegian

Norwegian is an SVO language, and as a result, the verb will generally precede the object (6).

- (6) Jeg *så* bilen.  
I saw car.the  
‘I saw the car’.

Furthermore, like its Germanic relatives (except English), it is also a V2 language. This means that the finite verbal element moves to the second position in all main clauses (typically analysed as verb movement to the CP domain, cf., e.g. Vikner, 1995). In this position, both finite auxiliaries and finite lexical verbs will precede not just the object, but also negation and other adverbs (7).

- (7) (a) Jeg *har ikke sett* bilen.  
       *I have not seen car.the*  
       'I haven't seen the car'.  
 (b) Jeg *så ikke* bilen.  
       *I saw not car.the*  
       'I didn't see the car'.

Moreover, interrogatives, illustrated by *yes/no*-questions in (8a,b) and *wh*-questions (8c,d), as well as in topicalised structures (9), V2 leads to inversion of the finite verb and the subject.<sup>1</sup>

- (8) (a) *Har du sett den?*  
       *Have you seen it*  
       'Have you seen it?'  
 (b) *Så du den?*  
       *Saw you it*  
       'Did you see it?'  
 (c) *Hva har du kjøpt?*  
       *What have you bought*  
       'What have you bought?'  
 (d) *Hva kjøpte du?*  
       *What bought you*  
       'What did you buy?'  
 (9) (a) *Idag har jeg kjøpt en bil.*  
       *Today have I bought a car*  
       'Today I have bought a car'.  
 (b) *Idag kjøpte jeg en bil.*  
       *Today bought I a car*  
       'Today I bought a car'.

### Residual Verb Second and Do-Support in English

Like Norwegian, English is an SVO language, as illustrated in (10).

- (10) I *saw* the car.

In contrast, however, English is not a V2 language but exhibits residual V2. This is a reflection of the fact that modern day English has remnants of a grammatical system that used to be more like the one observed in other Germanic languages today, where the finite verbal element typically was the second constituent in the clause. V2 in modern English is residual in two ways. While the other Germanic languages exhibit V2 in all clause types [cf. (6)–(9) above], V2 only applies in certain clause types in English. Moreover, while any finite verb has to move to the second position in Norwegian, V2 only applies to a subset of verbs in English, viz. auxiliaries. Consequently, only finite *auxiliaries* will precede negation and adverbs, as in (11a). In the absence of a finite auxiliary, the phenomenon of *do*-support emerges in negative declaratives.

- (11) (a) I *have not* seen the car.  
 (b) I *did not* see the car.

<sup>1</sup> In some dialects of Norwegian, including the dialect acquired by the three children in this study, inversion of the subject and the finite verbal element in *wh*-questions depends on type of *wh*-word, verb type, subject type and information structure. We will not specifically address this issue in the current paper, and we refer the reader to Westergaard (2009) for a discussion of this phenomenon.

Moreover, in *yes/no*-questions (12a,b) and *wh*-questions (12c,d), residual V2 leads to inversion of the finite auxiliary and the subject (12a,c). Again, in clauses without a finite auxiliary, *do*-support is required (12b,d):

- (12) (a) *Have you* seen it?  
 (b) *Did you* see it?  
 (c) What *have you* bought?  
 (d) What *did you* buy?

Finally, in topicalised structures, neither finite main verbs nor finite auxiliaries undergo movement across the subject in English. Furthermore, there is no requirement for *do*-support in the second position in these contexts:

- (13) (a) Today *I have* bought a car.  
 (b) Today *I bought* a car.

Note, however, that remnants of V2 can be found in certain topicalised structures, for example, clauses introduced by short adverbials such as *here* and *there*.<sup>2</sup>

Crucially, when such structures contain full DP subjects they trigger V2-like structures (14a, 15a), while with pronominal DP subjects they occur without V2 (14b, 15b).

- (14) (a) Here *comes the* bride.  
 (b) Here *she comes*.  
 (15) (a) There's *Noddy*.  
 (b) There *he is*.

Even though these structures are infrequent in English, they are relevant in this context because they provide evidence to the learner of a V2 grammar in English.

### Superficial Structural Overlap Between English and Norwegian and CLI

As mentioned in the 'Introduction' section, it has been argued that areas where the two languages in a bilingual situation display superficial structural overlap are particularly vulnerable to CLI. When looking at word order and verb placement in particular in English and Norwegian, there are several similarities. Both languages are SVO (16). Moreover, in negative declaratives, in *yes/no*-questions and in *wh*-questions the two languages display finite auxiliaries in parallel positions (17).

<sup>2</sup> This also applies to other types of topics. As shown in (i) and (ii), negative and restrictive elements typically trigger SAI:

- (i) Only then *did* he realise. . .  
 (ii) Never before *have* I seen a more beautiful. . .

Furthermore, verbs of reporting frequently invert when they follow direct speech, as in (iii). Note, however, that such verbs display a variation similar to that of *here* and *there* with regard to the placement of full DP and pronominal subjects [(iii)] versus [(iv)]:

- (iii) 'Where is he now?', *said* Mary.  
 (iv) 'Where is he now?', *she said*.

Finally, unaccusative verbs such as *arrive*, *sit* or *depart* may also precede the subject in topicalised structures, especially but not exclusively with locative elements [see (v) and (vi)]:

- (v) Then *arrived* the big stars.  
 (vi) In the corner *sat* a mysterious stranger.



- (16) Norwegian: Jeg så bilen.  
English: I *saw* the car
- (17) (a) Norwegian: Jeg *har ikke sett* bilen.  
English: I *have not seen* the car  
(b) Norwegian: *Har* du sett den?  
English: *have* you seen it?  
(c) Norwegian: Hva *har du kjøpt*?  
English: What *have you bought*?

In addition, English has verb movement of finite auxiliaries across the full DP subject in clauses introduced by adverbials such as *here/there*. This yields the same word order as in Norwegian:

- (18) Norwegian: Her *kommer bruden*.  
English: Here *comes the bride*.

However, when the finite verbal element is a lexical verb, rather than an auxiliary, the overlap breaks down. In Norwegian, lexical verbs also move to the second position in negative and interrogative clauses, while English employs *do*-support in these contexts (19).

- (19) (a) Norwegian: Jeg *så ikke* bilen.  
English: \*I *saw not the car*  
I *did not* see the car  
(b) Norwegian: *Så du* den?  
English: \**Saw you it?*  
*Did you* see it?  
(c) Norwegian: Hva *kjøpte du*?  
English: \**What bought you?*  
What *did you buy*?

In addition, the two languages show distinct patterns in non-subject-initial declaratives, where again, Norwegian has a consistent V2 pattern, while English has no verb movement to the second position (20):

- (20) (a) Norwegian: Idag *har jeg kjøpt* en bil.  
English: \**Today have I bought a car*.  
English: Today *I have bought* a car.  
(b) Norwegian: Idag *kjøpte jeg* en bil.  
English: \**Today bought I a car*.  
English: Today *I bought* a car.

We will argue that the superficial structural overlap shown above may lead to CLI.

## Previous Research on the Acquisition of Auxiliaries and *Do*-Support in English Monolingual Children

In this section, we first briefly address previous research on the acquisition of auxiliaries and *do*-support in English focusing on negative declaratives and interrogatives. Then we review some studies on verb placement in bilingual children acquiring English as one of their languages.

It is well known that children go through an early stage in which they systematically omit functional elements marking

tense and agreement. However, finite verbs are rarely completely absent from child grammars at this stage [commonly referred to as the Optional Infinitive (OI) stage, see, e.g. Harris and Wexler, 1996]. According to de Villiers and de Villiers (1985), auxiliaries enter English when children reach the two–three word stage. However, all auxiliaries do not come in simultaneously. Stromswold's (1990) extensive corpora study of 12 children (age range 1;2–7;10) shows that the first functional verbal element to appear is copula *be*, which on average is first attested at 2;2 in her data. A couple of months later, at 2;7, the first use of auxiliary *be* is found. The age of the first use of *do*-support is on average 2;8, while auxiliary *have* is the last, and only attested in Stromswold's data as late as at 3;5. See also Rispoli et al. (2012) for similar findings for copula *be*, *do* and auxiliary *be*. They do not discuss auxiliary *have*.

Most of the earlier studies on the acquisition of auxiliaries in negated and interrogative clauses have focused on auxiliaries other than *do*. One notable exception is Ervin-Tripp (1973) and Miller (1973), who describe *do*-support as typically first attested in negative declaratives, and subsequently expanded to questions. For four of the five children investigated in these studies, the productive use of *do* in negation preceded the use of *do* in questions by 2–7 months. The exception is Susan, who productively employs *do*-support in questions 2 months earlier than in negative declaratives. First attestations of *do* in both questions and negative declaratives is at age 2;2 for Susan. Fletcher's (1985) case study of Sophie finds a similar asymmetry where *do*-support is used in declarative clauses prior to questions.

For negative structures, the developmental path has been argued to involve an initial stage of pre-sentential negation, such as *No the sun shining* (Déprez and Pierce, 1993: 34, see also Bellugi, 1967 for an early description of this). These types of negative declaratives may occur with either *no* or *not*, and with or without the subject present. However, Drozd (1995) shows that only 10 of the 123 children investigated in his study produced at least one such structure, suggesting that not all children exhibit this behaviour. At the next developmental stage, children tend to produce structures with sentence medial *no* or *not* where the obligatory auxiliary typically is omitted, such as *Man no go in there* and *Wayne not eating it* (Radford, 1994: 152, 153). Radford refers to this as the (pre-functional) lexical-thematic stage, due to the fact that most main clauses are non-finite, most typically in the infinitive form. According to the original study in Bellugi (1967), children start using the negative forms *can't* and *don't* at this stage, but these represent unanalysed chunks, as auxiliaries generally tend to be absent. The frequent occurrence of non-agreeing *don't* has been related to the absence of adultlike tense and agreement at the OI stage (see, e.g. Schütze, 2010; Miller, 2013). At the final developmental stage, children rapidly start making use of auxiliaries in both negative and declarative contexts. This occurs at age 3;2 for Adam and 3;8 for Sarah, while Eve, who is widely considered to be very precocious, reaches this stage at age 2;2. Generally, these studies have not addressed whether there is a difference between the acquisition of *do* and other auxiliaries. However, Rowland and Theakston (2009) report a

lower proportion of target-like structures with *do*, compared to other auxiliaries, suggesting that *do* might be more difficult to acquire than auxiliaries in general. This is also true for the younger children in the study in Santelmann et al. (2002). In a recent study, Thornton and Rombough (2015) investigated the acquisition of *do*-support in negative declaratives in 25 children aged 2;5–3;4. They elicited negations where a target-like construction would include auxiliary *doesn't*. Their results show that more than half of the children's responses (52.5%<sup>3</sup>) are target-like and include *doesn't* (their Table 3). Only 10% of the responses contain just a bare main verb (*It not fit*). The most common non-target-like pattern involved non-target-like marking of third person singular (*It's not fit*, *It not fits*, *It doesn't fits*). However, the 25 children clearly split into two groups, one advanced group (12 children) and one less advanced group (13 children). The advanced group was target-like (using *doesn't*) 79% of the time, while the less advanced group only used target-like *doesn't* 1.4% of the time. In fact, nine of the 13 children in this latter group did not produce any instances of *doesn't* at all. The most common errors in this group involved either the pattern *It not V(s)* (33.1%) or non-agreeing *don't* [*It don't V(s)*] (17.2%). Notably, with respect to age, there does not seem to be any significant differences; both groups contain children within the whole age range from 2;5 to 3;4. This suggests that there is a lot of variation concerning at what age productive *do*-support in negative declaratives is acquired.

Turning to interrogatives, several studies have shown that auxiliaries tend to be omitted in *wh*-questions at an early stage (Roeper and Rohrbacher, 1994; Bromberg and Wexler, 1995). In a study on the acquisition of finiteness in English (and Norwegian) *wh*-questions, Westergaard and Bentzen (2010) investigate data from seven English-speaking children [Adam (3;0–3;5) and Sarah (2;9–5;1) from the Brown corpus, Brown, 1973; MacWhinney, 2000, and five children from the Manchester corpus, Warren, Anne, Ruth, Liz and Nicole ranging from 1;10–3;0, Theakston et al., 2001]. They report that copula *be* is much less frequently omitted compared to auxiliaries. Moreover, dummy-*do* and auxiliary *be* are missing much more often than modal auxiliaries. However, Westergaard and Bentzen (2010) do not find a clear distinction between the rate of dummy-*do* and auxiliary *be* omissions. Rather, there seems to be individual variation between the children with respect to which of the two auxiliary types are more frequently missing in *wh*-questions. Finally, their study also shows that *do* and auxiliary *be* are both still omitted quite frequently (for some children more than 50% of the time) up to the age of at least 2;9. In somewhat contrast to this, Erreich (1984) investigating 18 children aged 2;5–3;0 finds that auxiliaries are present in obligatory contexts in *wh*-questions and *yes/no*-questions (as well as declaratives) more than 80% of the time.

Concerning interrogatives, when auxiliaries are present in children's questions, SAI is typically employed. While some studies have reported that young children sometimes produce interrogatives without inversion (e.g. Klima and Bellugi, 1966 for

*wh*-questions, Erreich, 1984), Santelmann et al. (2002) point out that few studies have been able to show a stage that completely lacks SAI. Comparisons of the rate of SAI in *yes/no*-questions and *wh*-questions show variable results. Some studies do not find differences between the two types of interrogative clauses (e.g. Stromswold, 1990), others report that children more accurately and frequently make use of SAI in *yes/no*-questions than in *wh*-questions (Klima and Bellugi, 1966; Bellugi, 1971; Rowland, 2007; Pozzan and Valian, 2017), while yet others argue that SAI is employed earlier or more consistently in *wh*-questions than in *yes/no*-questions (Erreich, 1984; Valian et al., 1992).

Summing up, between ages 2 and 3 monolingual English-speaking children do not consistently include auxiliaries in negated and interrogative clauses, although inclusion of such elements gradually becomes the dominating pattern. The inclusion of dummy-*do* does not clearly lag behind the acquisition of other auxiliaries. Moreover, once auxiliaries are overtly expressed in negated and interrogative clauses, the typical patterns are Aux-Neg and SAI, although lack of SAI does occur in questions. Notably, to our knowledge, no studies report on non-target-like verb movement in monolingual English first language acquisition.

## Bilingual Children

The children in our study are acquiring English alongside the V2 language Norwegian, and we explore the effect this might have on the acquisition of verb placement in English. Although this has not been investigated for English/Norwegian bilingual children previously (though see Bentzen, 2000 for a preliminary study of one of the children in the current investigation), a few studies have looked at children acquiring English alongside other V2 languages. In an extensive case study, Knipschild (2007) investigates the acquisition of verb placement in the German/English bilingual boy Joshua, from age 2;4–3;1. While he appears to have acquired target-like V2 in German early on, he displays non-target-like behaviour in English. More specifically, he (predominantly at the earliest stages) produces structures that suggest verb movement of a lexical verb in negated and interrogative clauses (21), (22). He also employs verb movement in non-subject-initial declaratives (23). Furthermore, *do*-support only comes in after the age of 2;9, and is initially often used in non-target-like manners, e.g. uninverted (24a) or in declaratives as a superfluous *do* in non-emphatic contexts (24b) (from Knipschild, 2007: 92, 136):

- (21) I *want* not some water. (Joshua 2;4)
- (22) What *make* the kittens? (Joshua 2;10)
- (23) The flower *throw* daddy in the water. (Joshua 2;9)
- (24) (a) Where I *did* get this from? (Joshua 2;10)
- (b) I *did* watch it. (Joshua 2;10)

In fact, more than 90% of negated clauses and *wh*-questions displayed the patterns in (21) and (22) in the early stage (age 2;4–2;9). As pointed out in the previous section, monolingual English-speaking children hardly ever produce this kind of verb movement. Knipschild argues that the non-target-like utterances in (21)–(23) above are due to transfer from German. Similar findings in the English of bilingual German/English children have

<sup>3</sup> All percentages provided in the discussion of Thornton and Rombough's results are our own calculations.

been reported by Döpke (1998, 1999), Schelletter (2000) and Genske (2014).

In a case study of an Icelandic/English bilingual girl Katla, Bohnacker (2013) reports that while the child does use *do*-support, this is only employed in negative declaratives between the ages of 2;0 and 2;11 and in questions from the age of 3;0.

## Research Questions and Predictions for the Current Study

As highlighted in the 'Verb Placement in Norwegian and English' section, there is considerable superficial structural overlap between Norwegian and English, suggesting that CLI can be expected. Moreover, studies of bilingual children acquiring English together with other Germanic V2 languages reveal that such influence does occur. Given this, our research questions are as outlined in (1)–(3), and we make the predictions in (4) and (5).

- (1) To what extent is the acquisition of residual V2 and *do*-support in English affected by simultaneous acquisition of Norwegian?
  - (a) Is residual V2 expanded to apply to *all verb types*, including lexical verbs?
  - (b) Is residual V2 expanded to apply to *all clause types*, including topicalised structures?
  - (c) Is residual V2 expanded to both *all verb types* and *all clause types*, resulting in a full V2 system?
  - (d) Is the acquisition of residual V2 and especially *do*-support delayed or accelerated?
- (2) Are all the three children affected by CLI from Norwegian in the same way?
- (3) If they are not, what can explain the differences?
- (4) If CLI from V2 in Norwegian affects the acquisition of residual V2 and *do*-support in bilingual children, the following logical possibilities exist:
  - (a) If residual V2 is expanded to apply to *all verb types*, including lexical verbs:
    - (i) the acquisition of *do*-support should be delayed, as it makes the phenomenon superfluous in the grammar, and
    - (ii) lexical verbs should occur in the position normally reserved for auxiliaries in questions and negative declaratives.
  - (b) If residual V2 is expanded to apply to *all clause types*, SAI and *do*-support should also occur in topicalised structures.
  - (c) If residual V2 is expanded to apply to *all verb types* and *all clause types*, the children should allow a full V2 grammar.
  - (d) If CLI from Norwegian V2 accelerates the acquisition of residual V2 and *do*-support, these phenomena should be attested at an earlier stage in bilingual children.
- (5) Given that there is a great deal of ambiguity in the English system, it should be possible for different parsers to be affected by simultaneous input from Norwegian V2 in different ways, resulting in different grammars.
- (6) If such differences occur, they can be explained as an effect of language dominance.

## MATERIALS AND METHODS

The current study is a corpus study based on data from three girls, all bilingual from birth: Emma, Sunniva and Emily. These corpora were collected by the authors in connection with previous projects.<sup>4</sup>

As mentioned, the three girls grew up in very similar language situations; they all have one native English-speaking parent and one native Norwegian-speaking parent and grew up in Tromsø, Norway. Thus, English is a heritage language and Norwegian is the majority language in the lives of these children. The Norwegian-speaking parents opted to speak English with their children as well as with their English-speaking spouses. Thus, in all three cases, English is the home language. All girls attended nursery from around the age of one, Emily slightly later as she was born in the summer and started after the summer holiday, at approximately 14 months. Thus, this is the age at which consistent exposure to Norwegian started, even though both families were in close contact with family, friends and society at large, making some exposure to Norwegian likely most days even before the age of 1.

Two of the children, Emma and Sunniva, were also the first child in the family, while Emily has two older siblings, one of them being Sunniva. Emma's English-speaking parent is her American mother, while Sunniva and Emily's father is British. Sunniva and Emily also speak English with one another and with their brother. There is a 10-year age difference between the two sisters.

The data from Emma, Sunniva and Emily form the basis of the current study. Relevant information about the three corpora is summarised in **Table 1**. As **Table 1** shows, the corpora are quite spread out in terms of measures such as age, number of files and utterances, and Mean Length per Utterance for Words (MLU<sub>W</sub>). Emma was recorded biweekly in both English and Norwegian in the course of a three-month period between the ages of 2;7.10 and 2;10.9. There are six English files in Emma's corpus, consisting of 1831 child utterances. In these files, her MLU<sub>W</sub> range is 3.074–3.998. Sunniva was recorded in English and Norwegian for approximately a year, from age 1;6.25–2;8.0, at irregular intervals. There are nine files and 2512 utterances in her English data. Her MLU<sub>W</sub> ranges from 1.992–3.667 in these files. The equivalent information about Emma's and Sunniva's Norwegian files can be found in **Table 1** for comparison. Emily, on the other hand, was only recorded in English, and there are only four files in her corpus. The two first recordings were made just a few days apart, at ages 2;3.19 and 2;3.25, while recordings three and four were made considerably later and approximately

<sup>4</sup>We obtained written and informed consent from the children's parents on behalf of both their children's and their own participation in the corpus collection. At the time when these corpora were collected, there were no national requirements for approval of such data collection in Norway. However, the subsequent project which the current project is a part of, *Micro-variation in Multilingual Acquisition (MiMS)*, has gained approval from the Norwegian Centre for Research Data (NSD – <http://www.nsd.uib.no/nsd/english/index.html>).

**TABLE 1** | Overview of the data used in the study.

	Sunniva	Emma	Emily	Total
Age range	1;6.25–2;8.0	2;7.10–2;10.9	2;3.19–3;9.25	1;6.25–3;9.25
Number of English files	9	6	4	19
Utterances in English files	2512	1831	1495	5838
MLU range English files	1.992–3.667	3.074–3.998	2.833–4.961	1.992–4.961
Number of Norwegian files	7	7	Not applicable	14
Utterances in Norwegian files	2890	2222	Not applicable	5112
MLU range Norwegian files	1.932–3.442	3.282–4.120	Not applicable	1.932–4.120

a month apart, at 3;8.18 and 3;9.25. Her corpus consists of 1495 child utterances and the MLU<sub>W</sub> range is 2.833–4.961, with the first two recordings clustering between 2.8 and 3 and the last two ranging from 4.7 to almost 5.

Emma's files were originally transcribed by the Norwegian investigator and were later checked by a native speaker of English. Sunniva's and Emily's files were transcribed by a native speaker of English and subsequently checked by a native speaker of Norwegian. For the current study, the files from the three children were searched manually for the relevant structures. In the searches, all the contexts that obligatorily involve *do*-support or another auxiliary were identified: (i) negative declaratives, (ii) *yes/no*-questions and (iii) *wh*-questions. We included (iv) non-subject-initial declaratives in the searches, some of which also involve auxiliaries or *do*-support (see 'Residual Verb Second and *Do*-Support in English' and 'Superficial Structural Overlap Between English and Norwegian and CLI' sections).

Direct repetitions, both of other interlocutors and self-repetitions, were generally excluded. For example, the two examples in (25) were only counted as one *wh*-question. However, there are some exceptions to this. First, repetitions were included when the child kept repeating the same sentence but produced it in different forms [e.g. (26) below]. Similarly, when the child repeated an adult utterance incorrectly, the relevant example would be included in the count, but not when the child only repeated a part of the utterance (27).

- (25) Sunniva: Where's the pillow? (1;6.25)  
*Where's the pillow?*  
 Mother: Where's the pillow?  
 Mummy can't see the pillow from here.
- (26) Emily: Where are the cats gone? (3;9.25)  
*Where the cats gone?*  
*Where the cats gone?*  
 Mummy?  
 Mother: Hmm?  
 Emily: You have to say: Where are the cats gone?  
 Mother: Where have the cats gone?  
 Emily: Here.
- (27) Sunniva: What do you got? (2;1.16)  
 Mother: It's a little mouse, I think.  
*Sunniva: Little mouse, I think.*

Finally, identical repetitions produced by the child because she was explicitly asked to do so by the adult interlocutor were also included in the count. Other structures that were

excluded from the counts were utterances that were questioned in the transcription or for which alternative transcriptions were proposed, both indicating that the transcriber was unsure about the relevant utterance. Similarly, when a central part of the utterance is incomprehensible, the relevant example was not included in the count [see, e.g. (28)].

- (28) Mother: Mummy gonna put the (.) knickers on the little dolly. (1; 11.22)  
 Mother: On this little dolly.  
 Mother: I think they go on like this.  
*Sunniva: Where [?] xxx the knickers gone?*  
 Mother: Huh?

In other situations, examples where the incomprehensible part did not have any consequences for the relevant phenomena were included. Structures involving utterances where the transcriber was unsure about the transcription or where central parts of it were incomprehensible were double checked with the sound files and included or excluded depending on whether the authors agreed, disagreed or were still unsure about the relevant utterances.

## RESULTS

As **Table 2** shows, all three children productively use finite auxiliaries, modals and copula, and include these elements in negative declaratives and interrogatives quite consistently. Relevant examples are provided in (29)–(31). For Emily, non-target-like structures all lack an auxiliary [e.g. (32)], while Sunniva and Emma also have a couple of examples where the auxiliary is present but uninverted in questions [cf. (33)]<sup>5</sup>.

- (29) *Can you* sit on my back? (Emily 3;8.18)  
 (30) Winnie the Pooh's *not* broken mummy. (Sunniva, 2;6.1)  
 (31) What's *that here* on your watch? (Emma 2;8.17)  
 (32) I *not* doing it. (Emma 2;9.2)  
 TARGET: I'm not doing it.  
 (33) What *I'm* drawing, mummy? (Sunniva 2;6.1)  
 TARGET: What am I drawing?

<sup>5</sup>Given that the Tromsø dialect allows non-V2 in certain *wh*-questions, see footnote 1, structures such as (33) could be CLI from Norwegian. However, it is also possible that these are just examples of a type of behaviour that is sometimes observed in monolingual English children.



**TABLE 2 |** Target-like use of finite auxiliaries/copula in questions and negative declaratives versus non-target-like structures with missing auxiliaries or lack of SAI in Emma, Emily and Sunniva's files.

Child	SAI/Aux-neg (%)	No auxiliary (%)	No SAI in questions (%)	Total
Emma	90 (86.5%)	10 (9.6%)	4 (3.8%)	104
Emily	135 (95.1%)	7 (4.9%)	0	142
Sunniva	164 (93.2%)	9 (5.1%)	3 (1.7%)	176
Total	389 (92.2%)	26 (6.2%)	7 (1.7%)	422

For contexts requiring *do*-support, on the other hand, the situation is very different, especially for Emma. While 86.5% of her questions and negative structures include the auxiliary, *do*-support is only employed in 16.7% of structures requiring this. Furthermore, even though the majority of non-target-like utterances simply lack *do* (50%), similarly to what has been observed for monolingual English children, close to a third of Emma's questions and negative structures displays movement of a lexical verb (33.3%). In comparison, Emily and Sunniva supply *do* at a very similar rate to other auxiliaries, at 92.6 and 91.2%, respectively. These results are summarised in **Table 3**. Examples of target-like and non-target-like structures are provided in (34)–(39).

- (34) Where *did you* make [/] make some waffles?  
(Sunniva 1;11.22)
- (35) Cinderella, *do you* want to see? (Emily 3;9.25)
- (36) No, I *don't* watch mmm dance competition.  
(Sunniva 1;11.22)
- (37) I *hurt not* this knee now. (Emma 2;8.5)  
TARGET: I did not hurt this knee now
- (38) I *not know*. (Emma 2;8.17)  
TARGET: I don't know
- (39) *Drive daddy* me to barnehave? (Emma 2;8.5)  
TARGET: Will/did daddy drive me to nursery?

Considering these examples in more detail, we see that Emma in general is less target-like than Emily and Sunniva with all the structures involving residual V2. **Tables 4–6** provide the distribution of SAI for each of the structures requiring an auxiliary or *do*-support in the three children.

As these tables show, Emma is more target-like in *yes/no*-questions (78.5%) than in *wh*-questions (63.6%) and negative declaratives (47.7%). However, as shown in **Table 5**, the majority of Emma's *yes/no*-questions (47/65) involve an auxiliary other

than *do*, and these are all target-like. Of the 18 *yes/no*-questions requiring *do*-support, only 22.2% are target-like. All the non-target-like *yes/no*-questions involve movement of the lexical verb. As **Table 6** shows, most of Emma's non-target-like negative structures involve the omission of *do* (73.3%). Only a small proportion exhibit lexical verb movement (14.6%). The two other children are more consistent (and target-like) across the various structures.

Turning to non-subject-initial structures, we see that all three children are very different from one another. As shown in **Table 7**, Sunniva hardly produces any topicalised structures if we exclude topicalisations with *here/there* [cf. (13), (14) in the 'Residual Verb Second and *Do*-Support in English' section]. There are only two relevant examples attested in her corpus, and both of these are target-like. Both Emma and Emily produce topicalisations with verb movement. However, while Emma allows verb movement of both auxiliaries and lexical verbs, Emily only exhibits verb movement of auxiliaries. Furthermore, these two girls make use of verb movement at very different rates. As **Table 7** shows, Emma employs inversion in close to 30% of topicalised constructions, while Emily produces SAI at approximately 66%. Some examples are provided in (40)–(43).<sup>6</sup>

- (40) And *then did* Belle go up again. (Emily 3;8.19)
- (41) *Then you* need to take some shampoo. (Emily 3;8.19)
- (42) *Now throw* I it. (Emma 2;8.17)
- (43) *Now I* step on the scary woman. (Emma 2;8.17)

## DISCUSSION

So far, we have seen that both Emily and Sunniva are quite target-like in their use of auxiliaries in *wh*-questions, *yes/no*-questions and negation. With respect to *yes/no*-questions, Emma is also target-like in her use of auxiliaries, while she is somewhat less consistent in *wh*-questions and negative declaratives. However, when it comes to structures that require *do*-support, Emma is much less target-like than Emily and Sunniva. In this section, we discuss the results in more detail, and address the research questions posed in 'Research Questions and Predictions for the Current Study'. Recall that the main research questions concerned (1) to what extent the acquisition of residual V2 and *do*-support in Norwegian-English bilinguals is influenced by Norwegian V2, (2) whether all the three children are affected in the same way and (3) if they are affected differently, what can explain the differences.

## Different Outcomes of CLI

In what follows, we consider each of the three children in turn with regard to possible CLI from Norwegian. In doing so, we also consider whether the children are affected by this possible influence in the same way, or whether different parsers could possibly interpret the bilingual input in different ways. Recall that we predicted that CLI might take several

**TABLE 3 |** The total use of *do*-support in residual V2 contexts in Emma, Emily and Sunniva.

Child	<i>Do</i> -support (%)	No <i>do</i> (%)	V2 with lexical verbs (%)	Total
Emma	10 (16.7%)	30 (50%)	20 (33.3%)	60
Emily	63 (92.6%)	4 (5.9%)	1 (1.5%)	68
Sunniva	31 (91.2%)	2 (5.9%)	1 (2.9%)	34
Total	104 (64.2%)	36 (22.2%)	22 (13.6%)	162

<sup>6</sup>Note also that *all* topicalised elements were searched out, but very few these were arguments. The vast majority involved temporal and locative elements.

**TABLE 4 |** Finite verb placement in *wh*-questions (requiring SAI or *do*-support) in Emma, Emily and Sunniva's files.

Child	Auxiliaries/copula (%)			Do-support (%)			Total (%)	
	SAI	Lacking aux/cop	No SAI	SAI	Lacking <i>do</i>	V2/lex. verb	Target	Non-target
Emma	6/10	0/10	4/10	1/1	0/1	0/1	7/11	4/11
(11)	(60)	(0)	(40)	(100)	(0)	(0)	(63.6)	(36.4)
Emily	29/32	3/32	0/32	3/3	0/3	0/3	32/35	3/35
(35)	(90.6)	(9.4)	(0)	(100)	(0)	(0)	(91.4)	(8.6)
Sunniva	101/111	8/111	2/111	7/9	1/9	1/9	108/120	12/120
(120)	(91)	(7.2)	(1.8)	(77.8)	(11.1)	(11.1)	(90)	(10)

**TABLE 5 |** Finite verb placement in *yes/no*-questions (requiring SAI or *do*-support) in Emma, Emily and Sunniva's files.

Child	Auxiliaries/copula (%)			Do-support (%)			Total (%)	
	SAI	Lacking aux	No SAI	SAI	Lacking <i>do</i>	V2/lex. verb	Target	Non-target
Emma	47/47	0/47	0/47	4/18	0/18	14/18	51/65	14/65
(65)	(100)	(0)	(0)	(22.2)	(0)	(77.8)	(78.5)	(21.5)
Emily	53/53	0/53	0/53	16/20	4/20	0/20	69/73	4/73
(73)	(100)	(0)	(0)	(80)	(20)	(0)	(94.5)	(5.5)
Sunniva	54/56	1/56	1/56	10/11	1/11	0/11	64/67	3/67
(67)	(96.4)	(1.8)	(1.8)	(90.9)	(9.1)	(0)	(95.5)	(4.5)

**TABLE 6 |** Finite verb placement in negative declaratives (requiring an auxiliary or *do*-support) in Emma, Emily and Sunniva's files.

Child	Auxiliaries/copula (%)			Do-support (%)			Total (%)	
	Aux-neg	Lacking aux		Lacking <i>do</i>	V2/lex. verb		Target	Non-target
Emma	37/47	10/47	5/41	30/41	6/41	42/88	46/88	
(88)	(78.7)	(21.3)	(12.2)	(73.2)	(14.6)	(47.7)	(52.3)	
Emily	53/57	4/57	44/45	0/45	1/45	97/102	5/102	
(102)	(93)	(7)	(97.8)	(0)	(2.2)	(95.1)	(4.9)	
Sunniva	9/9	0/9	14/14	0/14	0/14	23/23	0/23	
(23)	(100)	(0)	(100)	(0)	(0)	(100)	(0)	

**TABLE 7 |** Finite verb placement in topicalised constructions, divided into verb types, in Emma, Emily and Sunniva's files.

Child	Lexical verbs		Auxiliaries/copula		Do-support		Total	
	Non-V2	*V2	Non-V2	*V2	Non-V2	*V2	Target non-V2	Non-target V2
Emma	22/28	6/28	25/39	14/39	1/1	0/1	48/68	20/68
(68)	(78.6%)	(21.4%)	(64.1%)	(35.9%)	(100%)	(0%)	(70.6%)	(29.4%)
Emily	16/16	0/16	7/33	26/33	3/27	24/27	26/76	50/76
(76)	(100%)	(0%)	(21.2%)	(78.8%)	(11.1%)	(88.9%)	(34.2%)	(65.8%)
Sunniva	1/1	0/1	1/1	0/1	0/0	0/0	2/2	0/2
(2)	(100%)	(0%)	(100%)	(0%)	(0%)	(0%)	(100%)	(0%)

forms. It might cause residual V2 to be expanded to include all verb types, which would result in lexical verbs preceding the subject in questions and non-subject-initial clauses and preceding negation in negative declaratives. A by-product of such influence could be that *do*-support is obsolete. A second possible outcome of CLI could be the expansion of residual V2 to all clause types, resulting in consistent SAI in topicalised declaratives. CLI might also affect both verb types and clause

types, resulting in (the possibility of) a full V2 grammar in the children's English. A final possibility is that the simultaneous acquisition of Norwegian V2 might accelerate the acquisition of residual V2 and especially *do*-support. The reasoning behind this prediction is that Norwegian V2 word order may enhance the need for a verbal element in a pre-negation position in negative declaratives and in a pre-subject position in interrogative clauses.

## Emma – Pattern 1: Transfer of Both Verb Types and Clause Types

As we saw in **Table 2**, auxiliaries and copula are acquired and occur in the target position in Emma's data 86.5% of the time (90/104). At the same time, **Table 3** shows that Emma is considerably less target-like in structures where *do*-support is required. As shown in **Table 6**, most non-target-like negative declaratives with a lexical verb are characterised by absence of *do*-support (73.2%). In addition, 14.6% of negative declaratives requiring *do*-support instead displays non-target-like movement of the lexical verb, as illustrated by the example in (44). Moreover, **Table 5** shows that as much as 77.8% of Emma's *yes/no*-questions involve non-target-like verb movement of the lexical verb, see (45) below:

(44) I hurt not this knee now. (Emma 2;8)

(45) Drive daddy me to barnehage (=daycare)? (Emma 2;7)

These examples indicate CLI from Norwegian V2 across verb types in Emma's data. Another indication of this is demonstrated by Emma's placement of *gonna* (not included in **Tables 3, 6**). *Gonna* (going to) is not a lexical verb but patterns with lexical verbs with regard to placement in negative declaratives and questions, as illustrated in (46). However, in Emma's negative declaratives, *gonna* almost exclusively occurs in front of the negation (19/22, 86.4%) (47):

(46) He's not gonna make it.

(47) Now that gonna not sleep more. (Emma 2;8)

Furthermore, there are also indications of CLI from Norwegian V2 across clause types in Emma's data. As revealed by **Table 7**, Emma allows verb movement of lexical verbs in non-subject-initial structures. 20/68 (29.4%) topicalisations display verb movement/inversion. Two examples are provided in (48) and (49), illustrating this for a lexical verb (48) and a perfective auxiliary (49).

(48) Now throw I it. (Emma 2;8)

(49) Now have I ringed Angus. (Emma 2;8)

Thus, Emma meets predictions (4a)–(4c). Due to influence from Norwegian V2, V2 appears to be transferred to apply across verb types and across clause types in Emma's English, making it almost equivalent to her Norwegian grammar in this respect.

## Emily – Pattern 2: Transfer of Residual V2 to Non-subject-Initial Clauses

As shown in **Table 2**, Emily includes auxiliaries and copula in the target position to a large extent in her residual V2 structures. However, unlike Emma, her use of *do*-support is also very target-like. Recall that she supplies auxiliaries in questions and negative structures in the target position at 95.1% (135/142) and *do* in the same structures at 92.6% (65/68). There is only one example of V2 with a lexical verb. Thus, it seems clear that Emily has not expanded residual V2 to apply to lexical verbs [thus not meeting prediction (4a)]. However, as illustrated in **Table 7**, Emily also exhibits SAI in topicalisations at 34.2% (26/76) and displays *do*-support in such contexts at 31.6% (24/76), as illustrated in (50)

and (51), making as much as 65.8% (50/76) of topicalised clauses non-target-like.

(50) And then was madame Gazelle's telephone ringing. (Emily 3;8)

(51) And then did Candy Cat come. (Emily 3;8)

Thus, we argue that there is CLI from Norwegian V2 into English also in Emily's data, causing residual V2 to be expanded across clause types, confirming prediction (4b).

## Sunniva – Pattern 3: Target-Like – And Early?

Finally, Sunniva is also target-like and consistent with respect to her use of auxiliaries and copula, including such elements 93.2% of the time (164/176), as shown in **Table 2**. She is also very target-like with *do*-support, which is included in 91.2% of required contexts (31/34) (see **Table 3**). However, note that there seem to be fewer contexts for *do*-support in Sunniva's files, compared to the other two, which might be related to the fact that Sunniva is younger than the other two children in most of her files. The general impression of Sunniva's production is that she is very target-like. She does not employ movement of lexical verbs (except in one instance), nor does she produce any non-target-like topicalisations [but notably she only produces two (non-imitated) non-subject-initial structures] (52).

(52) Maybe he's swimming. (Sunniva 1;9.13)

It would thus appear that Sunniva is not affected by CLI in residual V2 structures, contrary to predictions (4a)–(4c).

This leaves prediction (4d), suggesting that CLI from V2 may cause the acquisition of residual V2 and *do*-support to be accelerated. There are some challenges with respect to this issue. For one thing, the three children investigated in the current study are in relatively different age spans. Recall from **Table 1** that the children are recorded both at different age span and for different lengths of time. The distribution of recordings for the three children is presented in **Table 8**.

Furthermore, as discussed in the 'Previous Research on the Acquisition of Auxiliaries and Do-Support in English' section, *do*-support has been observed to occur in negation before questions in both monolingual and bilingual children. In our data, there is a total of 10 examples of *do*-support in all of Emma's files, five in negative structures, four in *yes/no*-questions and one in *wh*-questions. In Emily's first two files, aged 2;3.25 and 2;4.19, she produces 29 instances of *do*-support, four of which are in questions, which at least does not seem to be late compared to monolinguals. In Sunniva's files, however, there are 12 instances of *do*-support before the age of two; all but one occur in negative structures. Both types are illustrated in (53) and (54).

(53) No, I don't watch mmm dance competition. (Sunniva 1;11.22)

(54) Where did you make [/] make some waffles? (Sunniva 1;11.22)

**TABLE 8** | Overview of recordings according to age for Emma, Emily and Sunniva.

Age range	Sunniva	Emma	Emily
1;7–1;10	Sunniva (1;6.25) Sunniva (1;9.13)		
1;11–2;2	Sunniva (1;10.01) Sunniva (1;11.22) Sunniva (2;1.16) Sunniva (2;1.21)		
2;3–2;5	Sunniva (2;4.6)		Emily (2;3.19) Emily (2;3.25)
2;6–2;10	Sunniva (2;6.1) Sunniva (2;8.0)	Emma (2;7.14) Emma (2;8.5) Emma (2;8.17) Emma (2;9.2) Emma (2;9.23) Emma (2;10.8)	
2;11–3;1			
3;2–3;5			
3;6–3;9			Emily (3;8.18) Emily (3;9.25)

Even though Sunniva clearly makes use of *do*-support at a very young age, it is difficult to say for sure whether this is (i) early compared to monolinguals and (ii) early compared to Sunniva's general development.

With regard to the first question, Miller (1973), which is also based on corpus data, shows that Susan's first example of *do*-support in a negative clause is attested at age 2;0, and then there are several (11) at age 2;2. In questions, the first example is attested at age 2;2, and then two at age 2;3 and eleven at age 2;5. A quick search of three (randomly selected) children in the Manchester corpus (Theakston et al., 2001), Anne, Aran and Joel, reveals that the first attestation of *do*-support in the corpora of these children are at age 1;10.07, 2;0.09 and 1;11.01. In Sunniva's files, there are no examples of *do*-support at ages 1;6.25 and 1;9.13, and then the first example is attested at age 1;10.01 [*æ don't* (= I don't)]. Then the remaining eleven examples attested before the age of 2;0 are produced at 1;11.22. Thus, with regard to the question of first attestations, Sunniva does not seem that different from the three monolinguals from the Manchester corpus or Susan reported in Miller (1973). However, she does seem to make more extensive use of *do*-support than the three Manchester children at an early stage in development. In the files at ages 1;11.01, 1;11.29 and 2;0.26, which comprise a total of 2677 child utterances, Joel makes use of *do*-support in four instances (0.15%). Anne produces 10 examples among the 4339 utterances she produces at ages 1;10.07, 1;11.06 and 1;11.20 (0.23%). Finally, Aran makes use of *do*-support eight times among his 4139 utterances at ages 1;11.12, 2;0.09 and 2;1.07 (0.19%). In comparison, Sunniva's 12 examples are uttered in the course of the 1300 utterances she produces at ages 1;6.25, 1;9.13, 1;10.01 and 1;11.22 (0.92%). Thus, even though one from this cannot definitively conclude that she has acquired *do*-support earlier than these monolingual peers, it seems fair to say that

she uses *do*-support more than these monolingual children at this early stage of development.

Regarding the question of whether Sunniva's early use of *do*-support is simply a reflection of her generally being a precocious speaker, there are some indications that this might be the case. In the three early files in Joel's data (age 1;11.01–2;0.26), his  $MLU_W$  ranges from 1.299 to 1.846. Anne's  $MLU_W$  (aged 1;10.07–2;0.09) is between 1.558 and 2.233, while Aran's (1;11.12–2;1.07) is 1.299–2.341. In comparison, Sunniva's four early files (1;6.25–1;11.22) have  $MLU_W$  between 1.992 and 3.168, which is considerably higher than the three monolinguals, not just in absolute terms, but also in relation to her age. Sunniva's  $MLU_W$  is also high in comparison with the other two bilinguals, as illustrated in Table 9.

## Summary

The development of residual V2 and especially *do*-support can be shown to follow different paths in the three bilinguals in the current study. Emma exhibits CLI across verb types and across clause types, and thus shows a behaviour compatible with predictions (4a)–(4c). Emily transfers across clause types and makes use of residual V2 and *do*-support in non-subject-initial declaratives, confirming prediction (4b). The data from these two children put together show the validity of predictions (4b) and (4c). Emily's behaviour demonstrates that residual V2 may be expanded to apply to topicalised structures (4b), thus filling a gap in English in terms of how non-subject-initial structures pattern [cp. (8) and (9) to (12) and (13) in the 'Verb Placement in Norwegian and English' section]. Emma's grammar, however, expands residual V2 into general V2 both in terms of verb types and clause types and thus is in accordance with prediction (4c). However, based on these

**TABLE 9** | Overview of recordings according to MLU for Emma, Emily and Sunniva.

MLU range	Sunniva	Emma	Emily
1.9–2.1	Sunniva (1;6.25) Sunniva (1;9.13) Sunniva (1;10.01)		
2.2–2.4			
2.5–2.7			
2.8–3.0	Sunniva (2;1.16)		Emily (2;3.19) Emily (2;3.25)
3.1–3.3	Sunniva (1;11.22) Sunniva (2;1.21) Sunniva (2;4.6) Sunniva (2;6.1)	Emma (2;7.14)	
3.4–3.6		Emma (2;8.17) Emma (2;9.23) Emma (2;10.8)	
3.7–3.9	Sunniva (2;8.0)	Emma (2;9.2)	
4.0–4.2		Emma (2;8.5)	
4.3–4.5			
4.6–5.0			Emily (3;8.18) Emily (3;9.25)



results, we cannot be sure whether bilingual acquisition in this context could lead to CLI across verb types only (4a), resulting in a grammar where all verb types, including lexical verbs, may precede the negation in negative declaratives and subjects in questions, but not in topicalisations. Sunniva, on the other hand, does not exhibit any kind of transfer and is very precocious in her use of *do*-support. However, as she also appears to have a higher MLU<sub>W</sub> in relation to her age than both the three monolinguals we have compared her to and the other bilinguals in the current study, this is most likely not due to accelerated development as a result of CLI. Consequently, we cannot draw any conclusions regarding prediction (4d). However, these results together confirm prediction (5), as the simultaneous exposure to the V2 language Norwegian and residual V2 in English seems to result in different developmental paths for the acquisition of residual V2 and particularly *do*-support. Thus, it appears that different parsers may interpret the input differently in bilingual situations such as these. The question is what exactly might cause this to happen, specifically whether language dominance can explain the observed differences.

## Dominance as an Explanation for the Different Developmental Paths

The final research question addresses to what extent the differences between the three children can be explained with reference to language dominance, the underlying assumption being that CLI is more likely to occur from the dominant to the weaker language (for studies partially supporting this view, see, e.g. Bernardini and Schlyter, 2004; Nicoladis, 2006, 2012; Argyri and Sorace, 2007; Silva-Corvalán, 2014). As we have seen, Emma and Emily are affected by CLI, and consequently, we might expect Norwegian to be the dominant, or at least the stronger language for these girls. Sunniva, on the other hand, does not seem to be affected by CLI in the English structures under scrutiny, suggesting that her English is stronger, maybe even the dominant language for her. However, CLI is manifested in different ways in Emma and Emily, and another question pertains to whether these differences also can be explained by language dominance. Dominance has been argued to be an inherently gradient dimension (cf., e.g. Grosjean, 1982; Kupisch and Bernardini, 2007; Luk and Bialystok, 2013; Birdsong, 2015), and as such it should be possible for one of the girls to be more dominant in Norwegian than the other. If this were the case, the (inconsistent) whole-sale transfer of V2 observed in Emma's data would be indicative of a stronger Norwegian dominance for her compared to Emily, who displays a very specific transfer of residual V2 to topicalisations. However, recall from the discussion in the 'Materials and Methods' section, that we only have English data for Emily, so any comparison between the two will have to be made on the basis of English only.

The notion of dominance is ubiquitous in much of the literature on bilingualism, irrespective of whether the object of study is simultaneous or sequential bilingual child language acquisition, adult L2, or adult heritage speakers. In any bilingual situation, the question of which language is the stronger one

tends to be important and relevant. Despite this, a wide variety of measures have been used to determine language dominance, and there is consequently no generally agreed upon indicator available. The most frequently used measures relate to the (relative) level of proficiency in the two languages and/or the (relative) exposure to and use of the two languages (see, e.g. Kupisch and Bernardini, 2007; Silva-Corvalán and Treffers-Daller, 2015; Unsworth, 2015). Montrul (2015) includes all these three as different dimensions of dominance: the speaker's comparative proficiency, input situation and opportunity to use the languages. However, a recent paper (Lloyd-Smith et al., unpublished) introduces the experience-to-outcomes hypothesis to explain the wide range of variation usually observed in adult or adolescent heritage speakers, proposing that it is the sum of the speaker's experience in the heritage language that determines how proficient s/he becomes. This makes level of proficiency the result of the amount and quality of input and opportunities to use the language, rather than an interacting factor. In the end, the definition of dominance and the means used to measure it is to some extent dependent on the population investigated. For example, in adult or adolescent heritage speakers, amount of exposure and use may easily be operationalised as majority language, as the speakers clearly will have had more exposure to and opportunity to use this than the heritage language (see, e.g. Kupisch and van de Weijer, 2015). With young bilinguals, the situation is clearly different. Even though the three children in the current study will most likely end up with having had more exposure to Norwegian, this is not necessarily the case early on in the development, especially given their linguistic situations with English as their home language. Equivalently, even though they probably will end up more proficient in the majority language, this might not be the case at an early stage. Thus, in the current study, we discuss both language use and proficiency to determine to what extent dominance can explain the different behaviours of the three children. However, as the information available regarding language exposure and use is more limited, the main focus will be on proficiency.

There are no obvious objective measures available with respect to language exposure and use in the three small corpora. The families were not asked to fill in any questionnaires about language use, and the recordings were made so long ago that there is no reliable way of obtaining this information from the parents today. However, the language situations are very similar for the three girls. Recall from 'Materials and Methods' section that they all attended Norwegian nursery from approximately the age of one, and the families are strongly integrated with the community at large, thus ensuring exposure to and use of Norwegian from early on. With respect to the home language situation, all three girls grew up with one parent who is a native speaker of English, and one who is a native speaker of Norwegian, and they all have English as their home language. The non-native parents are highly proficient in English and do not make the kinds of mistakes that we have observed in two of the children. Indeed, these kinds of errors are lost early in the L2 acquisition of English by Norwegian learners (Westergaard, 2003). Moreover, two of the girls, Sunniva and Emily, even grew up in the same

family, making it less likely that huge variation in exposure to and use of English has caused the differences between them. If anything, Emily would have benefitted from the extra input from her older brother and sister. Furthermore, when she was born, the English grandparents were retired, which made it possible for them to visit their grandchildren more and for longer periods than when the older siblings were small. These facts together suggest that all the three children have a relatively balanced input situation, possibly slightly dominated by English at the earliest stage.

However, one possible explanation for the differences between the sisters might simply be that the data that we have available do not capture the period when Sunniva exhibits the same behaviour as Emily. Recall that Emily's non-target-like topicalised structures occur in the later files (at 3;8.18 and 3;9.25), while there are no non-subject-initial declaratives in Emily's early files (2;3.19 and 2;3.25).<sup>7</sup> Sunniva was recorded between the age of 1;6.25 and 2;8.0, and in this period, she only produces two (target-like) topicalised structures. Thus, Sunniva potentially may have gone through a period *after* data collection finished when she exhibited the same behaviour as Emily.

When proficiency has been used as an indicator of dominance, many different types of measures may be used to determine the balance between the two languages. MLU (sometimes with additional measures and/or specific implementations) is frequently been used in corpus studies (see, e.g. Genesee et al., 1995; Yip and Matthews, 2000, 2006; Bernardini and Schlyter, 2004; Kupisch and Bernardini, 2007; Hager and Müller, 2015). A general problem with the use of MLU to compare proficiency in the two languages of a bilingual is that languages differ greatly with regard to morphological complexity (Döpke, 1998; Yip and Matthews, 2006). This has been pointed out to be problematic for languages such as Italian and Swedish (Bernardini and Schlyter, 2004), as a comparison in terms of  $MLU_W$  underdetermines the score in Swedish compared to Italian because the Swedish definite article is suffixal, while the Italian one is a free morpheme. Consequently, only the latter would be included in an  $MLU_W$  count.<sup>8</sup> Note that the same difference applies between the two languages investigated here, Norwegian and English (*bil-en* versus *the car*). From this perspective, a higher  $MLU_W$  is expected in English, all other things being equal. Apart from this specific fact, the two languages are relatively similar with regard to morphological complexity and the realisation of various functional elements as free or bound.

An overview of the three children's  $MLU_W$  is provided in **Figure 1**. A visual comparison between Sunniva's  $MLU_W$  in English and Norwegian suggests that it is higher in English, especially between the age of 2;0 and 2;6. Emma's  $MLU_W$  appears to be more similar in the two languages, with English peaking in

one file and Norwegian in another. This impression is confirmed if we work out the average MLU for all the files in the each of the languages for Emma and Sunniva [inspired by Arencibia Guerra, 2008's measure of mean MLU difference (MMLUD), reported in Hager and Müller, 2015]. Emma's average  $MLU_W$  is exactly the same in the two languages (3.562), while Sunniva's average for English is 2.833 and for Norwegian 2.648. On this measure, both Emma and Sunniva would be classified as 'strongly balanced' according to Arencibia Guerra's (2008) criteria (there is less than a 0.29 difference between the languages). Recall, however, that  $MLU_W$  may to underdetermine the score for Norwegian compared to English because of the different status of definite article in the two languages. Nevertheless, with respect to  $MLU_W$ , both Emma and Sunniva are very balanced.

Another possible indicator of dominance is language mixing, as the direction of mixing often is claimed to occur from the stronger to the weaker language (cf. Genesee et al., 1995; Bernardini and Schlyter, 2004; Lanza, 2004; Kupisch, 2007; Kupisch and Bernardini, 2007, but see Anderssen and Bentzen, 2013). We only consider non-syntactic mixing in order to avoid that the instances of non-target-like verb movement investigated in the current study affects the measure of language dominance. Given previous studies, it is likely that the proportion of mixed utterances in the files of the three children (in both languages for Sunniva and Emma and English only for Emily) might give us an indication of which language the children is most proficient in. The language with the highest proportion of utterances with mixing should be the weaker, non-dominant one.

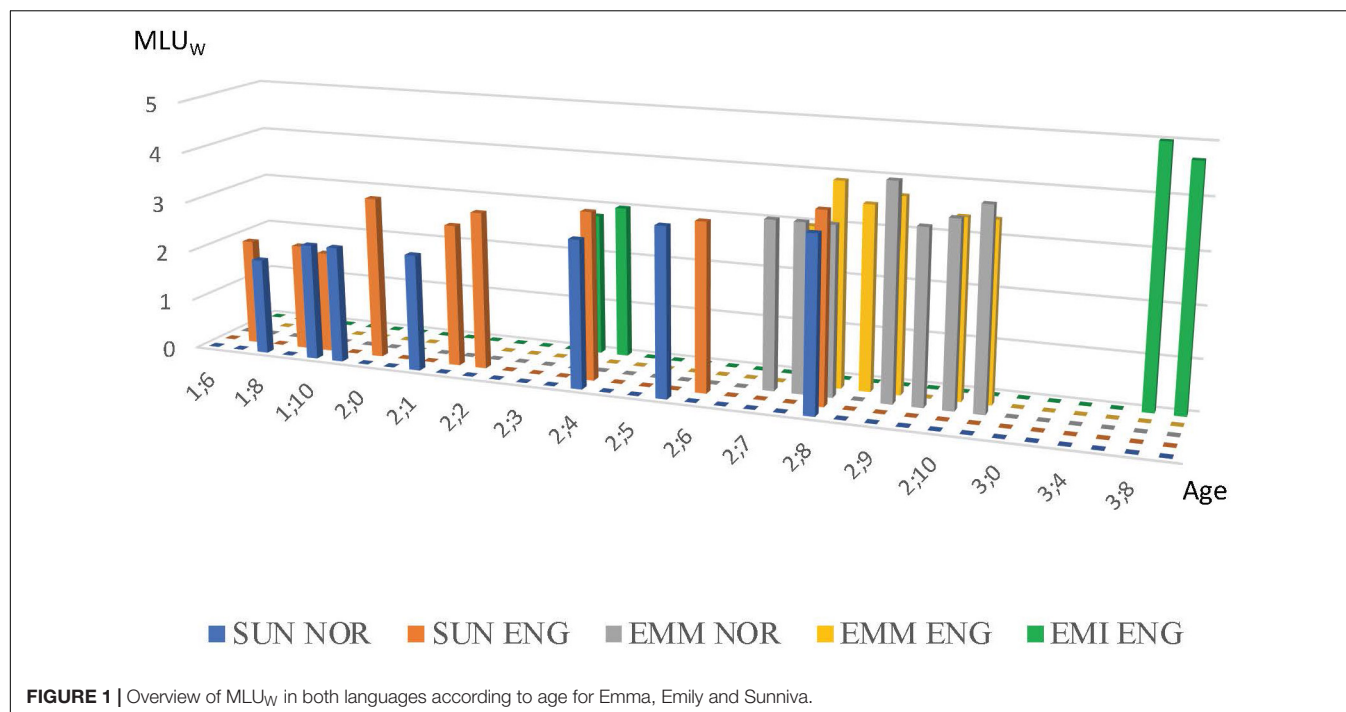
Correlating language mixing as a measure of language dominance and balance, we would expect both Emily and Emma to have a higher proportion of mixing in their English than Sunniva, possibly also with Emma mixing somewhat more than Emily (on the assumption that dominance can be gradient). However, as we can see from **Table 10**, this is not the case. In fact, Emma is the one with the lowest proportion of mixing in both languages, with 2.7% for English and 1.1% in Norwegian. Sunniva, who is the most target-like of the three children, mixes 8.2% in English and 4.9% in Norwegian. Finally, Emily, who we only have English files for, displays 13.4% mixes. Examples of the different kinds of mixing are provided in (55)–(57).

- (55) It wasn't *ordentlig* on *egentlig*. (Emily)  
TARGET: It wasn't *properly* on *really*.
- (56) *Nå skal dolly sove*. (Sunniva)  
TARGET: *Now the dolly is going to sleep*.
- (57) I falled on *en tå*. (Emma)  
TARGET: I fell on *a toe*.

These results thus reveal two things: (i) both Sunniva and Emma, the two children we have both English and Norwegian files from, mix more in their English than in the majority language and (ii) the proportion of mixing in the children's English does not seem to be correlated with the extent to which they behave target-like with residual V2 and *do*-support. Emma, the child who is the most influenced by Norwegian, mixes less than Sunniva, who appears to be the most target-like with respect to English verb placement. However, another factor also pertains

<sup>7</sup> There is one exception to this and that involves topicalised structures with *here* or *there* as topics, since they exhibit variable word order (see 'Residual Verb Second and *Do*-Support in English' section). These are also not included in Sunniva's data on non-subject-initial declaratives.

<sup>8</sup> As pointed out by a reviewer, Italian is a pro-drop language while Swedish is not. Thus, with respect to this feature,  $MLU_W$  should be higher in Swedish than Italian.



**TABLE 10 |** Language mixing with words, phrases and sentences in Emma, Emily and Sunniva.

Child/language	Total no. utterances	Word mixing	Phrasal mixing	Sentence mixing	Total mixing
Emma/ENG	1831	32 (1.7%)	8 (0.4%)	10 (0.5%)	50 (2.7%)
Emma/NOR	2222	14 (0.6%)	4 (0.2%)	6 (0.3%)	24 (1.1%)
Emily/ENG	1495	145 (9.7%)	35 (2.3%)	20 (1.3%)	200 (13.4%)
Emily/NOR			Not applicable		
Sunniva/ENG	2512	167 (6.6%)	19 (0.8%)	20 (0.8%)	206 (8.2%)
Sunniva/NOR	2890	123 (4.3%)	12 (0.4%)	6 (0.2%)	141 (4.9%)

to who the interlocutor is in the different files. In Emma's English files, she is mostly with her American mother, while in her Norwegian files, she is playing with an investigator whom she believes does not speak English. Sunniva's English files mostly include her Norwegian (but English-speaking) mother, while the Norwegian files are recorded with a Norwegian investigator. Note, however, that the mother is almost always present. Emily falls in between Sunniva and Emma in terms of acquisition of SAI and *do*-support but mixes the most of all of them. Emily's (English) files mainly include her Norwegian (but English-speaking) mother. One relevant question is why Emily mixes more than the other two. A factor contributing to this might be that Emily's older brother code-switches quite extensively, and in the context of a family where all the members are very fluent in both languages, code-switching is thus a natural communicative strategy also for Emily.

Further arguments against an explanation in terms of language dominance are provided in other studies involving Sunniva and Emma. Anderssen and Bentzen (2013) investigate modified definite DPs in Emma's Norwegian, finding overgeneralisation from English into Norwegian with respect to definiteness. They explain this behaviour with reference to simplicity, as these

structures involve so-called double definiteness in Norwegian. Importantly, this shows that CLI may also go from English into Norwegian in Emma's languages. Another study investigates the acquisition of gender in two monolingual Norwegian children as well as Emma and Sunniva (Rodina and Westergaard, 2013). With regard to this phenomenon, Sunniva and one of the monolingual children pattern together and are very target-like. Emma patterns with the other monolingual child, and both are non-target-like. This indicates that Sunniva is most likely more advanced than Emma in Norwegian as well.

To sum up, it appears that all the measures of dominance available to us indicate that Emma and Sunniva are fairly balanced bilinguals. Both of them have very similar MLU<sub>w</sub>s in English and Norwegian, and they mix more Norwegian into their English than the other way around. For Emily, we do not have access to Norwegian data, and thus cannot compare her English and her Norwegian competence. However, overall, she is not particularly delayed in her acquisition of English, which one might expect if Norwegian was strongly dominant. Furthermore, the rate of mixing does not seem to reflect the extent to which the three girls are target-like in their behaviour. Finally, the fact that

**TABLE 11 |** The use of + / –SAI in topicalisations with *here/there* adult and child speakers in the corpora.

Speaker	DP with SAI	Pronoun without SAI	Total <i>here/there</i>	Total adult input
EMMA mother	7 (58.3%)	5 (41.7%)	12	12
EMMA	17 (89.5%)	2 (10.5%)	19	
EMILY mother	24 (50%)	24 (50%)	50	52
EMILY sister	0	2 (100%)	2	
EMILY	19 (59.4%)	13 (40.6%)	32	
SUNNIVA mother	57 (55.3%)	46 (44.7%)	103	133
SUNNIVA father	15 (50%)	15 (50%)	30	
SUNNIVA	12 (60%)	8 (40%)	20	

English is their home language combined with frequent exposure to the community language supports the impression of three bilinguals who are very balanced. This also means that prediction (6) is not confirmed. The differences between the children cannot be explained with reference to language dominance. Rather, it seems that the three children (at least Sunniva and Emma) behave differently despite being relatively similar with regard to language balance. In the next section, we explore to what extent the different behaviours can be accounted for with reference to the two linguistic systems.

## Structural Ambiguity as an Explanation for CLI

So far, we have seen that the three bilingual children investigated in the current study behave very target-like when it comes to auxiliary placement in negative structures and questions. However, in contexts requiring *do*-support, the three children diverge. While Emily and Sunniva are very target-like also in these contexts, Emma produces a high proportion of non-target-like utterances. Recall from ‘Results’ section that unlike the other two, Emma employs verb movement of lexical verbs across the negation in negative structures and across the subject in *yes/no*-questions, suggesting that she has overgeneralised residual V2 to apply across verb types. Furthermore, she also allows both SAI and verb movement of lexical verbs in non-subject-initial declaratives, suggesting that she has overgeneralised V2 to apply across clause types as well, thus confirming predictions (4b) and (4c). For Emily, we have seen that

even though her behaviour is very target-like in structures involving residual V2, she overgeneralises auxiliary movement to topicalised structures, and as a result, the majority of her non-subject-initial declaratives involve non-target-like SAI or *do*-support. We have further seen that language dominance, at least as it can be measured with these data, cannot explain the differences between the children. Nevertheless, we observe that the parsers of the three children somehow interpret the data differently.

Recall from ‘Residual Verb Second and *Do*-Support in English’ section that even though English does not make use of SAI or verb movement in non-subject-initial declaratives, adverbials such as *here* and *there* in initial position may cooccur with V2 [cf. (14) and (15)]. One possible explanation for Emily and Emma’s non-target-like behaviour with topicalisations is that they have been exposed to a large number of these structures in the input, with a high proportion involving DP subjects, causing them to overgeneralise V2 into topicalisations in English. If this is the case, we would expect Emma and Emily to have had more exposure to and make more use of such structures than Sunniva. However, as illustrated by **Table 11**, both Emma and Emily appear to have considerably fewer of these topics in their input than Sunniva (12 and 52 versus 133). Also, while there is a slight majority of these structures with DP subjects, and hence V2, in the production of the adult speakers, the distribution is quite even. Similarly, the extent to which the children topicalise *here/there* is not completely in line with what the adult speakers in the same corpus do. Notably, Emma makes use of these topics more than her mother (19 versus 12), while Emily is the one with the highest number of these structures (32, compared to 52 by her mother and sister). The distribution of  $\pm$ V2 is also quite similar to that of the adults, but slightly more skewed towards V2. Emma is the one with the clearest preference for V2 in these structures (89.5%), but these are all target-like. It thus seems that the frequency of these structures in the input cannot account for the variation among the children (even though these corpora clearly are very limited).

Interestingly, however, an investigation into the children’s behaviour with DP versus pronominal subjects in non-subject-initial declaratives suggests that the two children who overgeneralise residual V2 are indeed influenced by the word order variation found with *here* and *there*. As demonstrated

**TABLE 12 |** Subject types and verb placement in topicalisations with *here/there* (for Emma, Emily and Sunniva) compared to other topics (for Emma and Emily only), divided into verb types.

Speaker	Here/there (%)		Other topics (%) lexical verbs		Other topics (%) Aux and be		Other topics (%) <i>do</i>	
	+V2/DP	–V2/Pr	+V2/DP	–V2/Pr	+V2/DP	–V2/Pr	+V2/DP	–V2/Pr
Emma	17/17 (100%)	2/2 (100%)	0/0 (0%)	22/28 (78.6%)	7/8 (87.5%)	24/31 (77.4%)	0/0 (0%)	1/1 (100%)
Emily	19/19 (100%)	13/13 (100%)	0/9 (0%)	7/7 (100%)	10/11 (90.9%)	6/22 (27.3%)	21/21 (100%)	3/6 (50%)
Sunniva	12/12 (100%)	8/8 (100%)	–	1/1 (100%)	–	1/1 (100%)	–	–



in **Table 12**, there is a strong tendency for both Emma and Emily to make use of V2 in exactly those cases where the subject is a DP, and not only when *here/there* are topicalised. The first two columns in the table show the children's use of V2 (+V2) with DP subjects and V3 (–V2) with pronouns when the topic is *here* or *there*. As the table reveals, the children follow this pattern completely. Then the next six columns show the same distribution (+V2/DP subject and –V2/pronominal subject) with other topics (e.g. *now*, *then* or *maybe*) and with lexical verbs, auxiliaries and copula *be* and *do*, respectively. Note that in these columns any structure that is +V2 is ungrammatical, but the closer the percentage in each of these columns is to 100%, the more similar the child's behaviour with topics in general is to *here/there*. As we can see, Emily consistently has V2 with DP subjects, except with lexical verbs, which she does not allow in the V2 position. With pronominal subjects, her behaviour is more variable, but clearly a substantial amount of her pronominal subjects also occurs with V2 (9/28, if we disregard lexical verbs), thus going against the pattern. Emma exhibits a high preference for both V2 with DP subjects (87.5%, but only auxiliaries are attested with DP subjects) and for V3 with pronominal ones (78.6, 77.4 and 100%). What is surprising is that it is the two children who appear to have been exposed to these structures the least who have adopted the word order pattern with V2. One possible explanation for this might be that the children need exposure to a certain number of examples to realise that structures such as non-subject-initial declaratives with *here/there* actually represent an exception. In the absence of sufficient exposure, the parser makes an overgeneralisation based on the available data, which for Norwegian/English bilinguals also will include data with massive indications of V2. Moreover, V-to-T movement of auxiliaries in English causes negative declaratives to superficially look similar to Norwegian constructions involving V2. The same is true for positive declaratives with adverbials such as *often* and *always* when they include an auxiliary. In the absence of auxiliaries, however, the similarity breaks down. This division between auxiliaries and lexical verbs makes the English system more ambiguous than the Norwegian one, leaving it open to several possible interpretations.<sup>9</sup> According to Henry and Tangney (1999: 139), 'language acquisition involves tension between the drive to create a maximally simple grammar in Universal Grammar (UG) terms and the need to adopt a grammar that covers the input data'; there is little doubt that the simultaneous exposure to English and Norwegian causes Norwegian to influence English residual V2. On the assumption that Henry and Tangney are correct, it is no surprise that CLI goes in this direction, as Norwegian V2, which is consistent across verb types and clause types, can be described as much more coherent than English residual V2. English is less consistent, with V2 only applying to certain structures (questions, negation and some topicalised structures) and specific verbs (auxiliaries, copula and in many cases *do*).

<sup>9</sup>See also Biberauer and Roberts (2017) for similar reasoning concerning how parameter setting may change diachronically.

## CONCLUSION

This paper investigates the acquisition of residual V2 in three Norwegian-English bilinguals. We find that the three girls exhibit three different patterns with regard to the relevant constructions, despite the fact that they grew up in comparable language situations. We argue that the non-target-like behaviour with respect to verb placement and *do*-support is caused by CLI from Norwegian. Furthermore, we have discussed various possible explanations for the differences between the three children's acquisition of verb placement in English. It is not obvious that the differences between the children can be explained with reference to language dominance, nor can they be explained in terms of frequency of exposure to non-subject-initial structures exhibiting optional V2. We have suggested that the observed CLI can be accounted for by the ambiguity in the English system, which leaves the data open to several possible interpretations when English is acquired in contact with the consistent V2 system in Norwegian. It thus seems that the children's parsers may interpret the input differently. Importantly, this means that the differences between the children are qualitative rather than quantitative. Furthermore, for Emily, we also know that she was able to 'recover' from this grammar, and we assume the same is true for Emma (who is an adult now), suggesting that this kind of recovery has to be possible and needs to be accounted for in developing grammars. We leave to future research the question of how such recovery from a non-target-like grammar is possible.

## ETHICS STATEMENT

This research is based on child language corpora collected in a period from 1999 to 2012. Parents were thoroughly informed about the data collection and the purpose of this, and signed consent forms on behalf of themselves and the children we collected data from. At the point in time when these data were collected, there were no requirements of approval from an ethics committee in Norway, and thus, such approval has not been obtained specifically for the collection of these data. However, approval has been obtained from NSD – Norwegian Centre for Research Data for the overall research project MiMS (Micro-Variation in Multilingual Acquisition and Attrition Situations) which the current investigation is a part of.

## AUTHOR CONTRIBUTIONS

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

## FUNDING

This project was partly funded by The Research Council of Norway (project 250857), and we gratefully acknowledge their support. We also want to thank the publication fund at UiT The Arctic University of Norway for funding the publication charges of this paper.

## REFERENCES

- Anderssen, M., and Bentzen, K. (2013). Cross-linguistic influence outside the syntaxpragmatics interface: a case study of the acquisition of definiteness. *Stud. Linguist.* 67, 82–100. doi: 10.1111/stul.12011
- Arencibia Guerra, L. (2008). Sprachdominanz bei bilingualen Kindern mit Deutsch und Französisch, Italienisch oder Spanisch als Erstsprachen [Language dominance in bilingual children with German and French, Italian or Spanish as their first language]. dissertation, Bergische Universität Wuppertal, Wuppertal.
- Argyri, E., and Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism* 10, 79–99. doi: 10.1017/S1366728906002835
- Bellugi, U. (1967). *The Acquisition of the System of Negation in Children's Speech*. dissertation, Harvard University, Harvard.
- Bellugi, U. (1971). "Simplification in children's language," in *Language Acquisition: Models and Methods*, eds R. Huxley and E. Ingham (New York, NY: Academic Press), 95–119.
- Bentzen, K. (2000). *I Like Not it Like du Like it': A Case Study of Language Transfer in Bilingual First Language Acquisition*. Master's thesis, University of Tromsø, Tromsø.
- 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
- Biberauer, T., and Roberts, I. (2017). "Parameter setting," in *The Cambridge Handbook of Historical Syntax*, eds A. Ledgeway and I. Roberts (Cambridge: Cambridge University Press), 134–162. doi: 10.1017/9781107279070.008
- Birdsong, D. (2015). "Dominance in bilingualism: foundations of measurement, with insights from the study of handedness," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 85–105.
- Bohnacker, U. (2013). "Reflections on dummy DO in child language and syntactic theory," in *Dummy Auxiliaries in First and Second Language Acquisition*, eds E. Blom, I. van de Craats, and J. Verhagen (Berlin: De Gruyter), 171–208.
- Bromberg, H., and Wexler, K. (1995). Null subjects in child wh-questions. *MIT Work. Papers Linguist.* 26, 221–247.
- Brown, R. (1973). *A First Language: The Early Stages*. Harvard, MA: Harvard University Press. doi: 10.4159/harvard.9780674732469
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Bristol: Multilingual Matters.
- de Villiers, J., and de Villiers, P. (1985). "The acquisition of english," in *The Crosslinguistic Study of Language Acquisition*, Vol. 1, ed. D. I. Slobin (Hillsdale, NJ: Lawrence Erlbaum), 27–139.
- Déprez, V., and Pierce, A. (1993). Negation and functional projections in early grammar. *Linguist. Inq.* 24, 25–67.
- 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
- Döpke, S. (1999). Cross-linguistic influences on the placement of negation and modal particles in simultaneous bilingualism. *Lang. Sci.* 21, 143–175. doi: 10.1016/S0388-0001(98)00021-7
- Droz, K. (1995). Child English: pre-sentential negation as metalinguistic exclamatory sentence negation. *J. Child Lang.* 22, 583–610. doi: 10.1017/S030500090000996X
- Erreich, A. (1984). Learning how to ask: patterns of inversion in yes-no and wh-questions. *J. Child Lang.* 1, 592–597. doi: 10.1017/S0305000900005961
- Ervin-Tripp, S. (1973). "Imitation and structural change in children's language," in *Studies of Child Language Development*, eds C. A. Ferguson and D. I. Slobin (New York, NY: Holt Rinehart and Winston), 391–406.
- Fletcher, P. (1985). *A Child's Learning of English*. Oxford: Basil Blackwell. doi: 10.1017/S1366728909990241
- Foroodi-Nejad F., and Paradis J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian-English bilinguals. *Bilingualism* 12, 411–427. doi: 10.1515/ling.1996.34.5.901
- Genesee, F., Nicoladis, E., and Paradis, J. (1995). Language differentiation in early bilingual development. *J. Child Lang.* 22, 611–631.
- Genske, K. (2014). *Early Child Bilingualism: Cross-Linguistic Influence in the Simultaneous Acquisition of German and English*. dissertation, Macquarie University Sydney/Universität Leipzig, Sydney, NSW.
- Grosjean, F. (1982). *Life with Two Languages: An Introduction to Bilingualism*, 11th Edn. Harvard, MA: Harvard College. doi: 10.1016/j.bandc.2009.07.008
- Hager, M., and Müller, N. (2015). Ultimate attainment in bilingual first language acquisition. *Lingua* 164, 289–308. doi: 10.1016/j.bandc.2009.07.008
- Harris, T., and Wexler, K. (1996). "The optional-infinite stage in child english: evidence form negation," in *Generative Perspectives on Language Acquisition*, ed. H. Clahsen (Amsterdam: John Benjamins), 1–42.
- Henry, A., and Tangney, D. (1999). "Functional categories and parameter setting in the second-language acquisition of Irish in early childhood," in *Language Creation and Language Change: Creolization, Diachrony, and Development*, ed. M. DeGraff (Cambridge, MA: MIT Press), 139–253. doi: 10.1017/S1366728900000353
- Hulk, A., and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism* 3, 227–244.
- Klima, E. S., and Bellugi, U. (1966). "Syntactic regularities in the speech of children," in *Psycholinguistic Papers*, eds J. Lyons and R. Wales (Edinburgh: Edinburgh University Press), 183–208.
- Knipschild, B. (2007). *Verb Placement, DO-insertion and Object Shift: A Case Study of a Bilingual German/English Child*. dissertation, University of Essex, Colchester. doi: 10.1017/S1366728906002823
- 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
- Kupisch, T., and Bernardini, P. (2007). "Determiner use in Italian-Swedish and Italian-German children: do Swedish and German represent the same parameter-setting?" in *Nordlys 34.3 Papers from the Language Acquisition Workshop, SCL 2006*, eds M. Anderssen and M. Westergaard, Tromsø, 209–229.
- Kupisch, T., and van de Weijer, J. (2015). "The role of the childhood environment for language dominance: a study of adult simultaneous bilingual speakers of German and French," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 174–194.
- Lanza, E. (2004). *Language Mixing in Infant Bilingualism: A Sociolinguistic Perspective*. New York, NY: Oxford University Press. doi: 10.1177/0142723711403980
- Liceras, J. M., Fernandes Fuertes, R., and Alba de la Fuente, A. (2011). Overt subject and copula omission in the Spanish and the English grammar of English-Spanish bilinguals: on the locus and directionality of interlinguistic influence. *First Lang.* 32, 88–115.
- Luk, G., and Bialystok, E. (2013). Bilingualism is not a categorical variable: interaction between language proficiency and usage. *J. Cogn. Psychol.* 25, 605–621. doi: 10.1080/20445911.2013.795574
- MacWhinney, B. (2000). *The CHILDES Project. Tools for Analyzing Talk. The Database*, 3rd Edn. Vol. II. Mahwah, NJ: Lawrence Erlbaum Associates. doi: 10.1080/10489223.2013.828061
- Miller, K. (2013). Variable input: what sarah reveals about non-agreeing don't and theories of root infinitives. *Lang. Acquis.* 20, 305–324. doi: 10.1080/10489223.2013.828061
- Miller, W.R. (1973). "The acquisition of grammatical rules by children," in *Studies of Child Language Development*, eds C. A. Ferguson and D. Slobin (New York, NY: Holt, Rinehart and Winston), 380–391.
- Montrul, S. (2015). "Dominance and proficiency in early and late bilingualism," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35. doi: 10.1017/S136672890500235X
- Müller, N., and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: italian and French as recipient languages. *Bilingualism* 4, 1–21. doi: 10.1017/S1366728911000101
- Nicoladis, E. (2006). Cross-linguistic transfer in adjective-noun strings by preschool bilingual children. *Bilingualism* 9, 15–32.
- Nicoladis, E. (2012). Cross-linguistic influence in French-English bilingual children's possessive constructions. *Bilingualism* 15, 320–328. doi: 10.1080/10489223.2016.1187615
- Patuto, M., Repetto, V., and Müller, N. (2011). "Delay and acceleration in bilingual first language acquisition: the same or different?" in *The Development of Grammar: Language Acquisition and Diachronic Change. In honour of Jürgen M. Meisel*, eds E. Rinke and T. Kupisch, (Amsterdam: John Benjamins), 231–261.

- Pozzan, L., and Valian, V. (2017). Asking questions in child English: evidence for early abstract representations. *Lang. Acquis.* 24, 209–233. doi: 10.1044/1092-4388(2011/10-0272)
- Radford, A. (1994). *Syntactic Theory and the Acquisition of English Syntax*. Oxford: Basil Blackwell.
- Rispoli, M., Hadley, P., and Holt, J. (2012). Sequence and system in the acquisition of tense and agreement. *J. Speech Lang. Hear. Res.* 55, 1007–1021. doi: 10.1044/1092-4388(2011/10-0272)
- Rizzi, L. (1996). “Residual verb second and the wh-criterion,” in *Parameters and Functional Heads*, eds A. Belletti and L. Rizzi (Oxford: Oxford University Press), 63–90.
- Rodina, Y., and Westergaard, M. (2013). The acquisition of gender and declension class in a non-transparent system: monolinguals and bilinguals. *Stud. Linguist.* 67, 47–67. doi: 10.1016/j.cognition.2006.05.011
- Roeper, T., and Rohrbacher, B. (1994). *Null Subjects in Early Child English and the Theory of Economy of Projection*. IRCS Report 94-16. Philadelphia, PA: University of Massachusetts and University of Pennsylvania. doi: 10.1044/1092-4388(2009/08-0037a)
- Rowland, C. F. (2007). Explaining errors in children’s questions. *Cognition* 104, 106–134. doi: 10.1016/j.cognition.2006.05.011
- Rowland, C. F., and Theakston, A. L. (2009). The acquisition of auxiliary syntax: a longitudinal elicitation study. Part 2: the modals and auxiliary DO. *J. Speech Lang. Hear. Res.* 52, 1471–1492. doi: 10.1044/1092-4388(2009/08-0037a)
- Santelmann, L. M., Berk, S., Austin, J., Somashekar, S., and Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: dissociating movement and inflection. *J. Child Lang.* 29, 813–842. doi: 10.1016/S0024-3841(03)00070-6
- Schelleter, C. (2000). “Negation as a cross linguistic structure in a German-English bilingual child,” in *Cross-linguistic Structures in Simultaneous Bilingualism*, ed. S. Döpke (Amsterdam: John Benjamins), 105–121. doi: 10.1080/10489223.2010.509270
- Schütze, C. T. (2010). The status of nonagreeing don’t and theories of root infinitives. *Lang. Acquis.* 17, 235–271. doi: 10.1080/10489223.2010.509270
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: determinants and mechanisms. *Linguist. Approaches Biling.* 3, 3–25. doi: 10.1017/S1366728904001610
- Serratrice, L., Sorace, A., Filiaci, F., and Baldo, M. (2009). Bilingual children’s sensitivity to specificity and genericity: evidence from metalinguistic awareness. *Bilingualism* 12, 1–19. doi: 10.1017/CBO9781139162531
- Serratrice, L., Sorace, A., and Paoli, S. (2004). Crosslinguistic influence at the syntax-pragmatics interface: subjects and objects in English-Italian bilingual and monolingual children. *Bilingualism* 7, 183–205.
- Silva-Corvalán, C. (2014). *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge: Cambridge University Press.
- Silva-Corvalán, C., and Treffers-Daller, J. (2015). “Digging into dominance: a closer look at language dominance in Bilinguals,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 1–14. doi: 10.1016/j.lingua.2008.09.008
- 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.
- Sorace, A., Serratrice, L., Filiaci, F., and Baldo, M. (2009). Discourse conditions on subject-pronoun realization: testing the intuitions of older bilingual children. *Lingua* 119, 460–477. doi: 10.1017/S0305000900004608
- Stromswold, K. (1990). *Learnability and the Acquisition of Auxiliaries*. dissertation, Massachusetts Institute of Technology dissertation, Cambridge: MA. doi: 10.1080/10489223.2014.943901
- Theakston, A. L., Lieven, E. V. M., Pine, J. M., 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.
- Thornton, R., and Rombough, K. (2015). The syntax-PF interface in children’s negative sentences. *Lang. Acquis.* 22, 132–157.
- Unsworth, S. (2015). “Amount of exposure as a proxy for dominance in bilingual language acquisition,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173.
- Valian, V., Lasser, I., and Mandelbaum, D. (1992). “Children’s early questions,” *Presented at the 17th Annual Boston University Conference on Language and Development*, Boston, MA. doi: 10.1075/eurosla.3.07wes
- Vikner, S. (1995). *Verb Movement and Expletive Subjects in the Germanic Languages*. New York, NY: Oxford University Press. doi: 10.1075/la.145
- Westergaard, M. (2003). Unlearning V2: transfer, markedness and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA yearbook* 3, 77–101.
- Westergaard, M. (2009). *The Acquisition of Word Order: Micro-Cues, Information Structure, and Economy*. Amsterdam: John Benjamins. doi: 10.1017/S136672890000033X
- Westergaard, M., and Bentzen, K. (2010). “Word order and finiteness in the acquisition of English and Norwegian wh-questions,” in *Proceedings of the 34th annual Boston University Conference on Language Development*, Vol 2, eds K. Franich, K.M. Iserman, and Keil, L (Somerville, MA: Cascadilla Press), 457–467.
- Yip, V., and Matthews, S. (2000). Syntactic transfer in a cantonese-english bilingual child. *Bilingualism* 3, 193–208.
- Yip, V., and Matthews, S. (2006). Assessing language dominance in bilingual acquisition: a case for mean length utterance differentials. *Lang. Assess. Q.* 3, 97–116. doi: 10.1207/s15434311laq0302-2

**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 CP and handling Editor declared their shared affiliation.

Copyright © 2018 Anderssen and Bentzen. 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.



# Multilingual Competence Influences Answering Strategies in Italian–German Speakers

Irene Caloi<sup>1\*</sup>, Adriana Belletti<sup>2,3</sup> and Cecilia Poletto<sup>1,4</sup>

<sup>1</sup> Institut für Romanische Sprachen und Literaturen, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Germany,

<sup>2</sup> Dipartimento di Scienze Sociali, Politiche e Cognitive, Università di Siena, Siena, Italy, <sup>3</sup> Département de Linguistique, Faculté des Lettres, Université de Genève, Geneva, Switzerland, <sup>4</sup> Dipartimento di Studi Linguistici e Letterari, Università degli Studi di Padova, Padova, Italy

## OPEN ACCESS

### Edited by:

Cornelia Hamann,  
University of Oldenburg, Germany

### Reviewed by:

Elisa Di Domenico,  
University for Foreigners Perugia, Italy  
Katja Francesca Cantone,  
Universität Duisburg-Essen, Germany

### \*Correspondence:

Irene Caloi  
caloi@em.uni-frankfurt.de

### Specialty section:

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

**Received:** 27 April 2018

**Accepted:** 26 September 2018

**Published:** 31 October 2018

### Citation:

Caloi I, Belletti A and Poletto C  
(2018) Multilingual Competence  
Influences Answering Strategies  
in Italian–German Speakers.  
Front. Psychol. 9:1971.  
doi: 10.3389/fpsyg.2018.01971

The present study aims at analyzing the role of nativeness, the amount of input in L1 acquisition and the multilingual competence in the performance of Italian–German bilingual speakers. We compare novel data from the performance of adult L2 learners (L1: Italian; late L2: German) and that of heritage speakers (heritage language: Italian; majority language: German) to previous data from monolingual speakers of Italian. The comparison deals with the produced word order at the syntax-discourse interface in sentences containing New Information Subjects in answers to questions that prompt the identification of the clausal subject. Overall, adult L2 speakers and heritage speakers perform alike but crucially differently from Italian monolinguals. These data reveal that multilingual proficiency determines an increased variety in the adopted answering strategies; in particular, the German-like strategy is active in Italian. Nativeness alone is thus no guarantee for a homogeneous performance across groups, nor do we find similar patterns of performance in speakers who grew up as monolinguals. Data also show heritage speakers' sensitivity to verb classes, with answering strategies varying in accordance with the verb argument structure. Participants' productions reveal an interesting relation in sentences with transitive verbs between subject position (pre-/postverbal) and object form (lexical DP/clitic pronoun).

**Keywords:** heritage language, L1 attrition, new information subjects, interfaces, optionality

## INTRODUCTION

This study addresses the issues of the role of nativeness, the amount of input in L1 acquisition and the multilingual competence in the performance of Italian–German bilingual speakers. We compare novel data from the performance of adult L2 speakers possibly undergoing attrition (AL2S; L1: Italian; late L2: German) and that of heritage speakers (HSs; heritage language: Italian; majority language: German) to previous data from monolingual speakers of Italian (MonoL1; Belletti and Leonini, 2004).

For the purpose of the present study, we rely on a concept of nativeness that corresponds to the exposure to the target language since birth in the familial environment, independently of the proficiency level ultimately attained later. As for the amount of input in L1 acquisition, we refer to the different linguistic settings in which the L1 is acquired. In a multilingual setting children



receive an input in the heritage language that is not as rich and differentiated as that of monolingual children. The third factor we address, i.e., multilingualism, takes into consideration the linguistic competence of our participants, who are advanced speakers of both Italian and German at the time of testing, independently of age of acquisition (bilingual child acquisition or adult L2 acquisition of German).

The three factors (although termed slightly differently) have been previously singled out by Montrul (2016, p. 17) as important variables in defining speakers' linguistic dominance. Along the lines of Kupisch and Weijer (2016, a.o.), we can refer to dominance as the strongest language in the speakers' competence. According to Montrul's model, proficiency is only one aspect of dominance, which usually correlates with other biographical and input variables (such as age of acquisition, place of birth, amount of input, type of context, etc.). Here, we will address the relationship between these variables and the attained proficiency. However, given that proficiency can vary depending on the linguistic level of analysis, we specify at the outset that we will focus on one specific phenomenon, rather than addressing a general linguistic assessment. The idea is to gain a better understanding of the role of single variables through the study of their reflex on a single linguistic phenomenon. This should ultimately contribute to highlight how dominance may be (re)set throughout the lifespan.

Our analysis addresses the produced VS and SV word orders determined at the syntax-discourse interface in sentences that are the reflex of a specific discourse content: the realization of New Information Subjects (NISs) in answers to questions that prompt the identification of the clausal subject.

We present the phenomenon referred to as *answering strategies* in Section "Answering Strategies." In Section "Subjects at the Interfaces: Previous Results From Multilingual Speakers" we discuss previous results from the performance of multilingual speakers in phenomena related to the syntax and interpretation of subjects at the interface with discourse in question-answer contexts. On this basis we will formulate our research questions in Section "Research Questions." In Section "Materials and Methods" we present the methods we used to collect the results presented in Section "Results." Section "Discussion" is dedicated to the discussion of the results. Section "Conclusion" gives the conclusions.

## ANSWERING STRATEGIES

In the present study, we consider a linguistic phenomenon that manifests itself at the interface between syntax and discourse, and we look at how a specific interpretive content is conveyed in the syntactic structure. Specifically, we are interested in NISs, which result from question-answer pairs aiming at identifying the clausal subject. The specific linguistic phenomenon is triggered by questions of the following kind:

- (1) *Chi ha vinto il premio?*  
who AUX win.PP the prize  
'Who won the prize?'

As exemplified, the question bears on the subject, which must be identified in the answer as the Focus of New Information, while the predicate and the object are presupposed in the given conversational context. As discussed in Belletti (2007) and in Belletti and Leonini (2004), languages differ in the way they answer questions that trigger NISs, i.e., they resort to different syntactic structures and different word orders to convey the intended meaning. Following the references quoted, we will refer to those selected structures as *answering strategies*.

Three main answering strategies have been identified in the languages investigated: verb-subject (VS) order with a postverbal subject; subject clefts, and subject-verb (SV) order with the preverbal subject bearing a characteristic prosodic prominence. The strategy preferably adopted in Italian exploits the VS order with a postverbal subject as in (2)A (Belletti, 2001, 2004, a.o.):

- (2) Q: *Chi ha vinto il premio?*  
who AUX win.PP the prize?  
A: *L' ha vinto Maria*  
obj.CL AUX win.PP M.  
Q: 'Who won the prize?' A: 'Maria won it'

In the same pragmatic context, French native speakers tend to produce subject clefts or reduced clefts (Belletti, 2007):

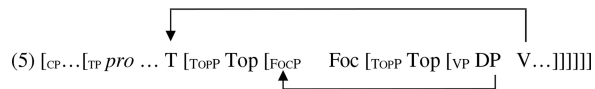
- (3) Q: *Qui a gagné le prix?*  
who AUX win.PP the prize?  
A: *C'est Marie (qui l' a gagné)*  
CL be M. who obj.CL AUX win.PP  
Q: 'Who won the prize?' A: 'It is Marie (who won it)'

In contrast, other languages like English and German tend to focalize the NIS in the preverbal position associating it to a marked prosodic prominence, as exemplified in the German example in (4):

- (4) Q: *Wer hat den Preis gewonnen?*  
who AUX the prize win.PP  
A: *MaRIE hat ihn gewonnen*  
M. AUX OBJ. ProN win.PP  
Q: 'Who won the prize?' A: 'Maria won it'

The three examples show how the same discourse function, i.e., the realization of NISs in answers to questions that aim at identifying the clausal subject, is carried out in different languages through strategies that differ in their syntactic structure and in their prosodic pattern. We assume with Belletti (2001, 2004, and subsequent work) the cartographic analysis according to which the low vP-peripheral area of the clause (TP) contains a discourse related Focus position dedicated to the New Information Focus interpretation and Topic positions along similar lines as the clause external Left periphery (Rizzi, 1997; Benincà and Poletto, 2004; Cruschina, 2009, 2012, and much subsequent work). As for the Italian example in (2)A, the structure is assumed to be obtained through the activation of the Focus position in the clause-internal vP-periphery, dedicated to host new information constituents, hence the NISs as well. According to this analysis, the low vP-peripheral position hosting the NIS is lower than the position targeted by the verb in its (head) movement within the

TP. Thus, the VS order is obtained with V moving over the low NIS along the lines schematically illustrated in (5):



In a null subject language like Italian, a silent referential *pro* is present in the high preverbal subject position, satisfying the relevant formal requirements (i.e., EPP, Chomsky, 1995; Cardinaletti, 2004; Rizzi and Shlonsky, 2007).

The structure of answering strategies with a postverbal subject ultimately relies on a number of syntactic mechanisms: the activation of the clause-internal low FocusP position for hosting the NIS, movement of V to T<sup>1</sup> and presence of *pro* in the preverbal subject position. It follows that the strategy cannot be exploited if the language is not a null subject language<sup>2</sup>. However, if some other way is available to satisfy the non-null subject property of the language, i.e., if the preverbal high subject position may be filled otherwise, for instance by an expletive subject pronoun, then the NIS can be left in the specifier of the low Focus position also in a non-null subject language, without violating any grammatical constraint. French cleft sentences of the type presented in (3) above illustrate one such case of how a non-null subject language can exploit the low vP-peripheral new information Focus position to express a NIS, essentially implementing a postverbal subject in disguise (Belletti, 2005, 2009, 2015; see also Hamann and Tuller, 2014, for an overview on French cleft and presentational constructions)<sup>3</sup>. In a nutshell (the relevant steps of) the derivation runs as follows. Let us assume that the core structure of clefts is built by a matrix clause containing the copula with its vP-peripheral discourse related projections including the New Information Focus one. The copula in turn takes as its complement a small clause (reduced) CP. The subject leaves its original position within the small clause CP and moves to the specifier of the low FocusP position in the vP-periphery of the copula in the matrix clause. Further movement of the copula to its functional T position results in the familiar VS word order. Finally, the high preverbal subject position is filled by the quasi-expletive pronoun *ce*. The basic features of the French subject-cleft answers are illustrated in (6):

- (6) [TP *Ce est*<sub>i</sub> [FOCP *Marie*<sub>j</sub> [VP <—<sub>i</sub>> [SM [PREDP [FinP *qui* [ <—<sub>j</sub>> *a gagné le prix* ] ] ] ] ] ] ]

<sup>1</sup>Or V- to the relevant non-finite past participial morphology as in examples like (2) containing a periphrastic Aux + Pst Prt tense. The head position targeted by the past participle is still higher than the discourse related vP-periphery.

<sup>2</sup>The availability of so called 'subject inversion' yielding the VS order and the null subject status of the language is indeed a core correlation in the classical literature on the null subject parameter (Rizzi, 1982; Jaeggli and Safir, 1989). The null subject property of the language is a necessary condition to allow for a VS order such as the one found in Italian type answering strategy. Notice, however, that nothing in principle rules out the grammatical possibility for a null subject language to also allow or even prefer a SV-type answering strategy in similar contexts. Bulgarian may precisely be a case in point according to the recent discussion in Genevsk-Hanke (2017).

<sup>3</sup>As discussed in the references quoted, many other diverse languages adopt the cleft strategy, among which, e.g., Norwegian, Malayalam, Japanese, Brazilian Portuguese.

Thus, Italian and French answering strategies to questions that trigger NISs, i.e., VS with a postverbal subject and subject clefts, both require the activation of the very same New Information Focus position in the vP-periphery. Cross-linguistic data ultimately offer robust evidence for the presence of such a position and its activation under the described discourse pragmatic conditions.

As mentioned above, at least one further strategy is attested cross-linguistically and that does not imply the activation of the low vP-peripheral Focus position, namely the focalization of the subject in preverbal position through prosodic prominence. The strategy is attested in Germanic languages, such as, e.g., English and German, but also in Romance languages such as Brazilian Portuguese (Dal Pozzo and Guesser, 2010) and, as recently discussed, in South American varieties of Spanish (Gabriel, 2010; Hoot, 2012; Leal et al., 2017), as well as in Bulgarian (Genevsk-Hanke, 2017; see footnote 2). The strategy consists in having the subject in preverbal position, yielding the SV linear order, and in attributing a characteristic prosodic prominence to the preverbal NIS.

In conclusion, given this brief summary, it clearly emerges that the shaping of an answering strategy is an articulated task, crucially involving both syntactic computations and their relation with the prosodic and interpretive interfaces<sup>4</sup>.

## SUBJECTS AT THE INTERFACES: PREVIOUS RESULTS FROM MULTILINGUAL SPEAKERS

Previous studies reported that multilingual speakers show optionality and non-target-like outputs for phenomena at the interface between syntax and pragmatics, such as, e.g., Topic shifts and NISs.

Sorace (2005, 2011) proposed the so-called Interface Hypothesis to provide a possible explanation for such results: phenomena that imply the integration of information from different cognitive systems may be more prone to instability in multilingual speakers ['unstable domains' in Sorace's (2011, p. 3) terms]. For instance, whereas Italian monolingual speakers agree in interpreting overt subject pronouns of subordinate clauses as Topic shift with respect to the main clause, thus selecting a referent different from the subject in the preceding matrix clause, English-Italian bilingual children (Sorace et al., 2009), attrited L1 speakers (Tsimpli et al., 2004), and advanced L2 learners of Italian (Belletti et al., 2007) show higher acceptance of coreference of the overt subject pronoun with the subject of the previous sentence, thus disregarding Topic shift<sup>5</sup>. Furthermore, the different types of bilingual/L2 speakers investigated may have access to a different possible grammatical analysis of the overt subject pronoun as a weak pronoun (in the sense of Cardinaletti

<sup>4</sup>See Belletti (2007) for the conclusion that the different answering strategies are in place from very early ages, based on a search on CHILDES (MacWhinney, 2000).

<sup>5</sup>Rinke and Flores (2018) report further cross-linguistic data on this issue: the authors claim that in European Portuguese the acquisition of correct interpretation of subjects takes longer for overt pronouns than for null pronouns both in monolingual and in (German-Portuguese) bilingual children.

and Starke, 1999; Cardinaletti, 2004), as the weak overt pronouns of their other language (e.g., she/he) are the equivalent of the Italian non-overt *pro*. Since the topic-continuity interpretation is available in their other language in the presence of an overt weak subject pronoun, bilingual/L2 speakers may overextend this interpretation also to overt Italian pronouns; however, in the same context monolinguals tend to prefer the weakest subject, i.e., the null variant, while overt subject pronouns tend to be interpreted as Topic-shift (Belletti et al., 2007, p. 672).

Given that answering strategies concern the production of subjects which express new information Focus at the discourse level, they also qualify as a potentially unstable domain in the linguistic performance of multilingual speakers. Belletti and Leonini (2004; see also Leonini and Belletti, 2004) tested which answering strategies adult learners of Italian produce when prompted to realize NISs. Their experimental group was not homogenous from the point of view of the participants' L1s and this clearly had a reflex on their different outputs, thus offering a straightforward interpretation of the results. Sixteen participants were German native speakers, who produced the target-like VS structure only in the 27% of the experimental items; while in the 68% of their answers they produced SV structures, with the focalized subject in the preverbal position. Three L2 learners with L1 French frequently produced answers in Italian with French-like clefts (69%) in order to express NISs<sup>6</sup>, in accordance to what has been described in Section "Answering Strategies" as the prototypical answering strategy in French. Thus, L2 speakers recognized the appropriate discourse context for the realization of NISs and reacted by activating an answering strategy which was not the one most prominent in the L2. Finally, Belletti and Leonini (2004) enrolled seven L2 learners with different L1s (e.g., Greek, Albanian, Polish); the group was very successful at producing target-like outputs of the VS kind (91%) across all verb types (range 87–93%). All participants in the third group were native speakers of null subject languages, which should explain their success at producing the target VS answering strategy<sup>7</sup>, in contrast with the German and French groups. As discussed in Section "Answering Strategies," the possibility to focalize the subject in the postverbal position crucially relies on the availability of a silent *pro* in the preverbal subject position, i.e., the null subject property, a necessary (although not sufficient, see footnote 2) condition for VS. The low production of VS structures shown by the French and German groups might be interpreted as a difficulty for learners to take into account all syntactic properties of *pro*, which finally results in the unsuccessful resetting of the null subject parameter. However, that this cannot be the case is indicated by the fact that null

subjects in Italian are largely available in both groups. Hence the conclusion must be that access to VS under the appropriate discourse conditions and availability of null subjects do not go together in L2 acquisition (see footnote 2 and the discussion in Belletti and Leonini, 2004). To sum up, results from this first study show that the L1 answering strategy remains active in adult L2 learners: when the production of NISs is elicited in L2 Italian, German speakers mainly produce SV answers, French speakers generally produce clefts, and only native speakers of null subject languages are successful in producing the target VS answer.

Postverbal subjects are hardly produced also by advanced learners who qualify as near-native speakers of Italian and have either British or American English as their L1 (Belletti et al., 2007). Results reveal that, despite their advanced acquisition of Italian, their use of the target-like VS strategy is still rather limited. Participants tend to produce SV structures with preverbal subjects focalized through prosodic prominence, in line with the dominant strategy in their L1<sup>8</sup>. Moreover, data from further independent tests run in the quoted study show that participants can use null subjects in an appropriate way<sup>9</sup>. This confirms the conclusion already drawn for non-advanced speakers of Italian reviewed above: most L2 speakers of Italian (with English-, German-, and French L1) show a dissociation between the use of null subjects and the production of postverbal subjects, with the latter not showing any significant development in time when the L1 is a non-null subject language. Overall, data from the two reviewed studies show a persisting difficulty in the production of target-like VS structures, with a parallel persisting activation of the prominent strategy of the native language (i.e., SV for English and German speakers).

In conclusion, the aspects of the background literature on the mastering of properties of subjects in Italian by multilingual speakers, relevant for the present study can be summarized as follows. Firstly, native speakers of Italian might show signs of attrition in this domain. The phenomenon, reported in Tsimpli et al. (2004) mentioned at the outset of this section, concerns Italian native speakers who qualify as near-native speakers of English and show altered interpretation of overt subjects. Whereas monolingual speakers of Italian interpret overt pronominal subjects of subordinate clauses as instantiations of Topic-shift with respect to the matrix clause subject, L2 speakers show a higher acceptance of coreference of the two subjects. Thus, attrition manifests itself in those speakers in the form of a broader acceptance of overt pronouns. As this work showed changes and attrition in the interpretation of subjects with respect to their overt/non-overt pronominal realization, this further encourages us to investigate whether another discourse-related property, i.e., the pre-/postverbal position of the overt subject might similarly undergo attrition.

Despite correct use of the null subject property, L2 learners of Italian show persisting difficulties at achieving a target-like use of the related property yielding the order VS in answers

<sup>6</sup>The phenomenon is even more striking when data analysis takes into consideration the argument structure of the verbs in use: cleft production by L1 speakers of French is particularly high in elicited Italian answers containing either a transitive verb (88%) or an intransitive verb of the unergative kind (80%). Differences among verb types will be further addressed in Section "Results."

<sup>7</sup>These data are confirmed also by an independent study run with L1 speakers of Polish: Labuz (2012) ran the Polish version of the test in use in Belletti and Leonini (2004) with 16 monolingual L1 speakers of Polish and found that they produced VS structures in 84% of the answers with NISs. The strategy remains active in Polish speakers of L2 Italian (15 participants), who perform almost target-like in Italian (95% of VS answers).

<sup>8</sup>Except for sentences that include existential structures of the *c'è/ci sono* kind (i.e., 'there is/there are'), as was also the case for the non-advanced French and German groups discussed earlier.

<sup>9</sup>Although overproduction of overt pronominal subjects can also be detected in their oral production (Belletti et al., 2007, p. 672).

containing a NIS even at advanced stages of acquisition. The fact that target answers are correctly produced in the second language only by those speakers whose L1 is a null-subject language suggests that the successful acquisition of the relevant answering strategy might be dependent on an early setting of the parameter in child acquisition. In this line of reasoning, we speculate on the idea that the native answering strategies cannot be easily inhibited in L2, especially in the case in which these lead to the production of grammatical sentences in the target language (although infelicitous in the given context as, e.g., use of SV instead of VS in Italian).

Results from the two groups, i.e., attrited native speakers and L2 speakers, seem to point to two different hypotheses for the mechanism that shapes answering strategies. On the one hand, L1 attrition within the domain of pronominal subject interpretation shows that this is actually an unstable domain, whose system can be influenced by advanced L2 acquisition. In this vein, we can hypothesize that multilingualism shapes answering strategies in forms that depend on the properties of the languages involved. On the other hand, the observation of L2 learners suggests the hypothesis that answering strategies are crucially shaped in childhood, and L2 learners experience difficulties in inhibiting the native strategy when this offers grammatical options in the L2 (as it is the case for subject clefts and SV structures in Italian). We can therefore hypothesize that answering strategies are in place from very early ages and eventually keep stable despite the presence of competing L2 grammatical options, thus turning nativeness and amount of input in L1 acquisition into the crucial factors for shaping answering strategies. If this hypothesis is correct, we should find an Italian native-like performance in our multilingual speakers. We take into consideration both hypotheses in what follows by analyzing the role of multilingualism, nativeness and amount of input in L1 acquisition in answering strategies.

## RESEARCH QUESTIONS

In the previous section, we sketched out two plausible routes for the shaping of answering strategies in multilingual speakers. In order to enlarge the body of data at our disposal and to analyze the two hypotheses, we run a test on the production of answering strategies in two groups of multilingual speakers who share the same languages. All our participants are Italian–German speakers; however, for one group of speakers, Italian is the native language and German is the L2; for the second group of participants, Italian is the heritage language and German is the majority language. We refer to the first group as adult L2 speakers (AL2S), and to the second group as HSs. To the best of our knowledge, these two categories with the Italian–German combination have never been tested before for the production of answering strategies. Together with data from previous studies on L2 acquisition (Belletti and Leonini, 2004; Leonini and Belletti, 2004) and on near-native speakers of Italian (Belletti et al., 2007), we will complete the picture of how answering strategies are computed by Italian–German multilingual speakers. In particular, by choosing participants who speak Italian and German in different

settings and with different acquisition histories (adult L2 acquisition of German, or Italian as heritage language), we want to investigate how different factors contribute in shaping answering strategies. In particular, the factors that we take into consideration are nativeness, amount of input in childhood, and multilingualism.

As pointed out by Kupisch and Rothman (2016, a.o.), the opposition between HSs and native speakers is not correct from the theoretical point of view; in fact, HSs are also native speakers of the heritage language. In other words, they cannot be opposed to native speakers, because they are native speakers of the heritage language themselves. From this point of view, HS, AL2S and monolingual speakers do not differ, because they all started acquiring Italian from birth in the familial environment, such that they can all be considered native speakers of Italian. If being a native speaker is the crucial factor in shaping answering strategies, we should find homogeneous performances across speakers' profiles. However, one important difference among the groups might be the amount of input received during L1 acquisition. Monolingual speakers and AL2S all grew up only with one language; therefore, we can assume that they all received a comparable amount of input in the critical period. In contrast, HSs grew up in a community characterized by a majority language other than the heritage language. Hence, the input they received in the heritage language is different from that received from the two groups who grew up in a monolingual setting in Italy, both from the quantitative and the qualitative point of view. Although this is subject to extreme individual variation, HSs have on an average a more limited access to the heritage language because the majority language usually covers some relevant communicative functions (e.g., education, interaction with peers in public spaces, TV shows, etc.). As for the linguistic phenomenon at stake, HS are exposed to both answering strategies (VS and SV answers) in the input through the two languages. By comparing the three groups of speakers, i.e., monolingual speakers of Italian and adult L2 speakers (of German, L1 Italian) on one hand, and HSs (of Italian, German the majority language) on the other hand, we aim at verifying the role played by the input during language acquisition in shaping answering strategies.

Moreover, subtle differences between AL2S and HS might emerge in the two groups as an effect of different syntactic conditions. As briefly mentioned in Section “Subjects at the Interfaces: Previous Results From Multilingual Speakers,” previous studies already reported that the argument structure of the verb in use could have an influence on the adopted answering strategy (see Belletti and Leonini, 2004; Belletti et al., 2007). The analysis of the collected data will take into consideration the verb class as a relevant factor in order to draw a comparison between the experimental groups. Differences among verb classes might reveal further interesting aspects of the performance of multilingual speakers.

For the same reason, we want to add one further research question, which concerns the realization of objects in transitive structures. Although the study focuses on the production of NISs, structures containing transitive verbs also offer the opportunity to observe how the internal argument is produced in the specific



discourse context. In fact, the object is usually given in the question in use for triggering the answering strategy with NISs, such that the element is a Topic and should be realized as a pronoun in the answer, specifically as a clitic pronoun in a language with clitics like Italian. We will observe which strategy multilingual speakers put in use to convey Topic-like objects and their interaction with NISs.

To sum up, in the present study we intend to address the following research questions:

- (i) Do multilingual speakers produce different answering strategies with respect to monolingual speakers of Italian?
- (ii) If so, does the type of acquisition setting (monolingual/bilingual) play any role in shaping answering strategies? To what extent do the strategies for NISs realization differ in the two groups?
- (iii) Does the verb class affect the type of answering strategy activated by participants?
- (iv) How are objects realized when the answer contains a transitive verb?

With respect to questions (i) and (ii), we can make the following speculations: if nativeness is the decisive factor in shaping answering strategies, we expect AL2S and HS to perform similarly to MonoL1 speakers. Alternatively, it could also be argued that the decisive factor is the amount of input received during L1 acquisition; if that is the case, we expect AL2S and MonoL1 to perform similarly, but not HS. Finally, if multilingualism, and the subsequent presence of both conflicting strategies in the input through the two languages, is the decisive factor, we should find a third pattern of performance: HS and AL2S performances should pattern alike, but crucially differ from that of MonoL1. The pattern of performance across the three experimental groups (MonoL1, AL2S, and HS) should reveal which factor among nativeness, amount of input in L1 acquisition, and multilingualism plays the bigger role in shaping answering strategies. Moreover, results could shed light on the kind of input received by HS in the multilingual environment they grew up in, since native speakers with an advanced command of the L2 (and therefore potential attrition of the L1) ultimately represent the privileged source of input for HS.

## MATERIALS AND METHODS

### Participants

The following groups of speakers took part in the present study: 22 adult heritage speakers of Italian (HS) with German as the majority language, and 20 adult L2 speakers (AL2S), who are native speakers of Italian and acquired German as L2 after childhood. Their results will also be compared to those of a group of monolingual speakers of Italian (MonoL1), whose data are reported from the original study presented in Belletti and Leonini (2004).

Heritage speakers grew up in Germany and were exposed to Italian since birth by either one or both parents, who were native

speakers of Italian<sup>10</sup> and used the language in the daily interaction with the child (see Rothman, 2009). At the time of testing all HS reported to use Italian as the family language together with German. Except for two speakers whose parents had moved away from Italy during childhood, all other HS participants were second generation children of parents who left Italy in their early adulthood. HS participants were educated in the majority language, but 20 of them (out of 22) also took formal courses of Italian at school and/or at the university. Based on these data, we assume that HS have received a reduced amount of input in the heritage language with respect to monolinguals<sup>11</sup>. At the time of testing, all participants were enrolled as undergraduate students at the Goethe University of Frankfurt. No minimum level of proficiency in Italian was required to take part in the study. However, in choosing participants for this study we particularly took into consideration two factors: formal education and contact with the family of origin. First, in line with the studies discussed in Kupisch and Rothman (2016), we assume that formal education in the heritage language and literacy can allow for higher proficiency and closer to monolingual-like performance. Second, the young age of the participants translates into closer and on-going relationships to their families of origin, which plays a relevant role in heritage language maintenance. As previously pointed out by O'Grady et al. (2011) and by Polinsky (2011), heritage language attrition can take place over the life-span as an effect of reduced contacts to the family and the community of origin, i.e., reduced use of the heritage language in everyday life. We aimed at recruiting high performing participants by choosing young adults who are (or have been) engaged in formal courses of Italian on a regular basis and who maintain a tight contact with the family of origin in which Italian is spoken. We expect the enrolled participants to have benefitted from their exposure to multiple diversified speakers in a number of varieties and registers during their education and from the contacts with the family of origin and, possibly, with the Italian community. Although the two factors will not be analyzed as experimental factors in our analysis, we point out that, based on previous results from the literature (O'Grady et al., 2011; Polinsky, 2011, a.o.; Kupisch and Rothman, 2016), they were included in the guidelines for the recruitment of high-performing participants.

Adult L2 speakers (AL2S) were native speakers of Italian, who grew up in Italy as monolingual speakers of Italian, moved to Germany as young adults (most AL2S moved to Germany during University years), and learnt German as L2. In most cases, relocation to Germany was preceded by some formal courses of German as L2 taken at Italian schools or universities. Participants from this group all reported using German on a daily basis both at work and at home, although in variable amounts. The same holds true for Italian: they all reported to use Italian daily, in

<sup>10</sup>In 10 families both parents were native speakers of Italian; in 10 families one parent was a native speaker of Italian and one parent was a native speaker of German; in 2 families one parent was a native speaker of Italian and the other parent was a native speaker of a further minority language (i.e., Hungarian or Polish), but this was not transmitted to the child.

<sup>11</sup>For reasons of space we cannot discuss here how the amount of input in the target languages can be measured in a bilingual context, but see Grüter and Paradis (2014) for detailed discussions.

familial interactions and/or in a variety of entertaining and social activities (e.g., watching/reading the news, films, books, social media interaction, etc.). Although to the writers' knowledge there is no acknowledged minimum amount of exposure time to the L2 in order to allow for the onset of attrition effects, we nonetheless set the requirement of a minimum of 4 years of continuative stay in Germany with the described systematic use of German for AL2S participants to enroll in the study.

The characteristics of HS and AL2S participants are summarized in **Table 1**, together with the information on the monolingual speakers, whose data are reported from Belletti and Leonini (2004).

As reported in **Table 1**, AL2S were older than HS on average: this is intrinsically due to the characteristics of the participants we decided to enroll. On the one hand, AL2S learnt German as L2 after childhood, moved to Germany as young adults, and certainly needed some years in order to achieve a good competence of German. On the other hand, HS are young undergraduate students, who still live with the family of origin in most cases. This explains why it would have not been possible to match the two participant groups for age.

In the fourth column, we report data on the amount of years spent in Germany by participants in the two groups. On average, AL2S speakers had spent 13;7 years in Germany at the time of testing (range: 4–28;3 years). As for HS, the data roughly corresponds to their age, as they spent all their lives in Germany, except for medium-term periods of stay in Italy (less than 1 year<sup>12</sup>).

Both groups underwent two cloze tests, one for Italian and one for German: both versions of the test consisted in a text from a newspaper article<sup>13</sup>, from which several functional and lexical words were erased. Participants were requested to fill in the gaps. Despite the absence of at ceiling performances, AL2S completed the Italian version of the test better than the German one, whereas HS showed the opposite pattern, with better performance in the German than in the Italian test. Results from the two groups differ in the Italian test (interval of accuracy in AL2S = 0.80–0.90; and in HS = 0.49–0.63), but not in the German one (interval of accuracy in AL2S = 0.57–0.80; and in HS = 0.74–0.81). We take the results of AL2S on the German test as a proof of their good command of the L2.

As to MonoL1 participants, Belletti and Leonini (2004) tested 10 native speakers of Italian, who came from different Italian regions. Their age ranged between 20 and 33 years old. No cloze

test was performed in the original study as their command of Italian was evident.

## Materials

We collected data on answering strategies through the elicited production task first presented in Belletti and Leonini (2004).

Participants watched 22 short videos and listened to 40 experimental questions, which triggered answers with NISs. Videos depicted characters involved in daily activities and ended with one of the actors asking a question on the identification of the subject. One or two further questions were also audio played at the end of each video concerning the event represented in the scene to serve as distractors. Participants were instructed to produce oral answers expressing the verb (thus allowing for the observation of the subject position); they were also explicitly encouraged to answer the questions in the way that sounded the most natural to them.

Experimental questions are distributed across four conditions, i.e., 20 finite sentences containing a transitive verb (7), 4 sentences with an unaccusative verb (8), 10 sentences with an unergative verb (9), and 6 sentences featuring existential structures with the Italian copula (10).

- (7) *Chi ha aperto la finestra?*  
who AUX open.PP the window  
'Who opened the window?'
- (8) *Chi è arrivato?*  
who AUX arrive.PP  
'Who arrived?'
- (9) *Chi ha urlato?*  
who AUX scream.PP  
'Who screamed?'
- (10) *Cosa c'è sopra il tavolo?*  
what CL AUX on the table  
'What is on the table?'

The test material also included 19 fillers in the form of questions that concerned the video content but did not trigger answers with NISs. Experimental questions and fillers were randomized throughout the task.

## Procedure and Coding

Participants were tested individually and the test took approximately 12–15 minutes. Their outputs in the elicited production test were recorded with an aLLreLi (ALLCP0033\_Q9G) digital voice recorder and later transcribed and coded by a researcher. Data from the cloze tests described above were coded by two students: a native Italian speaker

<sup>12</sup>One participants also reported to have lived for 1 year in an English-speaking country during childhood.

<sup>13</sup>The two texts in use differ for the two languages (one was not the translation of the other), such that the comprehension of one text could not bias the completion of the other one.

**TABLE 1** | Participants' characteristics (data for MonoL1 are reported from Belletti and Leonini, 2004).

Participants	N°	Age range Mean (SD)	Years in Germany (SD)	Cloze-t IT (SD)	Cloze-t GE (SD)
AL2S	20	26–60 41;1 (8;3)	13;7 (5;1)	85.5% (2)	69% (5)
HS	22	21–38 25;6 (5)	25;6 (5)	56.4% (3)	77% (1)
MonoL1	10	20–33	–	–	–

coded the data from the Italian cloze test, while a native German speaker coded the data from the German cloze test. All results were cross-checked by a second researcher.

Since the goal of our study is the observation of the position of NISs, only answers containing (at least) the subject and the verb were relevant for our analysis. Therefore, all subject-only answers were discarded (11). In addition, non-relevant answers were excluded from our analysis; i.e., answers that did not contain the required Subject of New Information (12–13).

- (11) *La ragazza*  
the girl  
'The girl'
- (12) *Non ho visto*  
not AUX see.PP  
'I didn't see it'
- (13) *Non lo so*  
not CL know  
'I don't know'

Only main clauses containing the subject and the verb were analyzed and classified under either one of the three following categories, depending on the word order and the syntactic structure: VS answers (14), SV answers (15), and Other answers (16–17). The label Other was used for any grammatical sentence, whose syntactic structure did not correspond to the SV or VS word order. It turned out that the majority of them were (reduced) clefts (16) or passive structures (17).

- (14) *L' ha aperta la ragazza*  
obj.CL AUX open.PP the girl  
'The girl opened it'
- (15) *La ragazza l' ha aperta*  
the girl Obj.CL AUX open.PP  
'The girl opened it'
- (16) *Era la ragazza*  
was the girl  
'It was the girl'
- (17) *La finestra è stata aperta dalla ragazza*  
the window AUX.be.pass open.PP by-the girl  
'The window was open by the girl'

Answers with transitive verbs were also further analyzed depending on whether the clausal object was produced as a clitic pronoun [see example (14–15) above] or as a lexical DP, as in (18):

- (18) *La ragazza ha aperto la finestra*  
the girl AUX open.PP the window  
'The girl opened the window'

One further aspect we took into consideration with transitive verbs is the position of the object with respect to the verb and the subject (e.g., SVO or VOS); the issue will be addressed in details in the "Results" section.

## RESULTS

Outputs from participants in the AL2S and HS groups were transformed into percentages according to the answering strategy in use (VS/SV/Other) in order to allow for a comparison with results from the MonoL1 group (reported from Belletti and Leonini, 2004). **Table 2** offers a descriptive overview of the results from the three groups.

The data analysis was carried out using SPSS Statistics Version 17.0. Pairwise comparisons were run in order to analyze the outputs of the three experimental groups. Data revealed that both AL2S and HS perform differently from MonoL1 in the production of VS and SV answers. First, we compared AL2S against MonoL1. The Mann-Whitney tests indicated that the number of VS answers is higher for MonoL1 speakers ( $Mdn = 98.5$ ) than for AL2S speakers ( $Mdn = 66.66$ ),  $U = 3.00$ ;  $p < 0.001$ ;  $r = 0.78$ . In turn, more SV answers were produced by AL2S ( $Mdn = 11.36$ ) than by MonoL1 ( $Mdn = 0$ ), Mann-Whitney  $U = 17.50$ ,  $p < 0.001$ ,  $r = 0.67$ . No significant difference was found in the production of Other answers between MonoL1 ( $Mdn = 0$ ) and AL2S ( $Mdn = 3.03$ ), Mann-Whitney  $U = 67.00$ ,  $p = 0.155$ ,  $r = 0.28$ . Second, we compared HS against MonoL1. The Mann-Whitney tests indicated that the number of VS answers is higher in MonoL1 ( $Mdn = 98.5$ ) than in HS ( $Mdn = 51.25$ ),  $U = 5.000$ ,  $p < 0.001$ ,  $r = 0.75$ . The SV answering strategy was more frequent in HS participants ( $Mdn = 45.97$ ) than in MonoL1 speakers ( $Mdn = 0$ ), Mann-Whitney  $U = 6.500$ ,  $p < 0.001$ ,  $r = 0.74$ . The two groups did not differ in the production of Other answers (MonoL1  $Mdn = 0$ , HS  $Mdn = 0$ , Mann-Whitney  $U = 92.000$ ,  $r = 0.14$ ). Based on these results, we conclude that both AL2S and HS perform differently from MonoL1 when producing elicited answering strategies with NISs.

We also compared AL2S and HS and found that the latter produce more SV (HS  $Mdn = 45.97$ ) than the former (AL2S  $Mdn = 11.36$ ), Mann-Whitney  $U = 137.00$ ,  $p = 0.037$ ,  $r = 0.32$ . As for the production of VS and Other answers, no significant difference was found between the two groups of multilingual speakers.

**Table 3** shows how participants' outputs distribute across answering strategies with respect to the kind of verb included in the question-answer pairs, i.e., transitive verbs, unergative verbs, unaccusative verbs, and existential structures.

**TABLE 2 |** Production of VS/SV/Other answers by MonoL1/AL2S/HS (in percentages, SD, median).

	MonoL1	AL2S	HS
VS	98	61.59	50.52
SD	2.87	26.31	27.95
Median	98.5	66.66	51.25
SV	1.0	26.03	44.75
SD	2.16	27.91	28.78
Median	0	11.36	45.97
Other	1.0	12.38	4.73
SD	1.54	19.88	8.81
Median	0	3.03	0

**TABLE 3 |** Production of VS/SV/other answers by MonoL1/AL2S/HS in sentences with transitive, unergative, unaccusative verbs, and existential structures (in percentages).

	MonoL1			AL2S			HS		
	VS	SV	Other	VS	SV	Other	VS	SV	Other
Transitive	97.00	2.00	1.00	50.00	30.34	19.66	36.82	55.81	7.37
Unergative	99.00	1.00	–	66.49	26.63	6.88	49.77	46.37	3.86
Unaccusative	95.00	5.00	–	70.00	30.00	–	60.23	39.77	–
Existential	96.00	–	4.00	99.00	1.00	–	91.51	5.31	3.18

Since MonoL1 speakers performed very consistently (see **Table 3**) across syntactic conditions (VS range 95–99%; SV range 0–5%, Other range 0–4%), we are going to set this group apart for a moment in order to focus our analysis on multilingual speakers (AL2S and HS).

Mann–Whitney tests revealed that the use of SV answers significantly differs between AL2S and HS in sentences with transitive verbs and with unergative verbs. As for transitive verbs, the Mann–Whitney test showed that the production of SV answers is greater in HS participants ( $Mdn = 60.0$ ) than in AL2S participants ( $Mdn = 11.1$ ),  $U = 129.00$ ,  $p = 0.022$ ,  $r = 0.35$ . HS participants also produce more SV answers than AL2S speakers in sentences with unergative verbs (HS  $Mdn = 47.2$ , AL2S  $Mdn = 12.5$ , Mann–Whitney  $U = 349.50$ ,  $p = 0.041$ ,  $r = 0.31$ ). Finally, no significant difference between the two groups is attested when the elicited answers contain an unaccusative verb or an existential structure. The two groups (AL2S and HS) produce comparable numbers of VS and Other answers across all verb types (no significant difference revealed by Mann–Whitney tests).

We also wanted to look at data from a different perspective in order to analyze the distribution of VS and SV answers across conditions, and to check for relations between verb types and sentence structures. Our intention was to verify whether the argument structure of the predicate in use in the question–answer pairs plays a role in determining how NISs are produced, with respect to the overall activated answering strategy and in particular to the subject position.

It is evident from the data presented above that a strong relation holds for at least one condition, namely the one with the Italian copula for existential structures: in this condition, participants from both the AL2S group and the HS group are very consistent in replicating the VS word order, with the NIS following the copula (AL2S  $mean = 99.00\%$ ;  $SD = 4.35$ ,  $Mdn = 100$ ; HS  $mean = 91.51\%$ ;  $SD = 22.9$ ;  $Mdn = 100$ ). The pattern is very robust and alternative strategies are attested very infrequently. As for the remaining conditions, i.e., transitives, unergatives, and unaccusatives, no such straightforward result is observable.

Kruskal–Wallis tests did not reveal any significant difference within AL2S speakers in the production of VS and SV answers with transitive, unergative and unaccusative verbs. Based on this observation, we conclude that verb type does not play a role in determining which answering strategy is adopted by AL2S (not counting existential structures).

In contrast, data from the HS group reveal one interesting property: although the number of SV answers is stable across conditions (no significant difference revealed by Kruskal–Wallis tests), that of VS answers is not (see **Table 4**). From the descriptive point of view, the number of VS structures is at its lowest alongside transitive verbs (36.82%), it increases with unergatives (49.77%) and becomes significantly higher with unaccusative verbs (60.23%). The Kruskal–Wallis test revealed that there was a difference between the number of VS answers produced with different verb types not quite reaching significance [ $H(2) = 5.72$ ,  $p = 0.057$ ] with a mean rank of 40.32 for unaccusatives, 33.66 for unergatives and 26.52 for transitives. Therefore, the argument structure of the verb in use in the sentence seems to have an influence on whether HS speakers adopt the VS answering strategy.

Based on the lowest number of VS answers reported alongside transitive verbs, we deduce that the presence of two arguments in the sentence might represent a relevant factor in determining the adopted structures. For this reason, we run a third round of analysis on answers with transitive verbs and observe how objects are realized. Our analysis takes into consideration two factors: (a) the object form, i.e., whether it is produced as a clitic or as a full-fledged lexical DP, and (b) its position with respect to the subject and the verb. As a result, different possible structures are attested for VS answers as well as for SV answers.

Starting with the first factor, i.e., the object form, data reveal that participants from both groups produce objects both as clitics and as lexical DPs (see **Table 5**).

This observation is particularly interesting in consideration of the fact that lexical DPs were not expected in this context, yet they

**TABLE 4 |** Production of VS answers by HS participants in sentences with transitive, unergative, unaccusative verbs (in percentages, SD, median).

HS speakers	Transitive	Unergative	Unaccusative
VS mean	36.82	49.77	60.23
SD	32.7	30.37	33.13
Median	34.1	52.8	70.85

**TABLE 5 |** Object production analysis (N° and in percentages).

	DP	Clitic
AL2S	192/343 (56.0%)	151/343 (44.0%)
HS	162/320 (50.6%)	158/320 (49.4%)



characterize half of the answers. In the syntactic and pragmatic context offered by the experimental conditions in combination with the videos, objects of transitive verbs always appear in the eliciting questions and are therefore Topics, which express given information in the answer. The repetition of the object Topic as a full lexical noun phrase in the answer is not felicitous; this is characteristically the condition in which the usage of a clitic pronoun is required, as indeed the behavior of native speakers from previous studies confirms (Leonini and Belletti, 2004; Belletti and Rizzi, 2017).

As for the second factor, i.e., the position of the object within the sentence structure, the analysis cannot be carried out without taking into account the subject position and therefore the overall answering strategy in use. In what follows we focus on VS and SV answers in turn and analyze the attested word order in the two strategies<sup>14</sup>. In VS answers the following word orders are attested: Clitic–Verb–Subject (cVS in 19), Object–Verb–Subject (OVS in 20), Verb–Object–Subject (VOS in 21), and Verb–Subject answers with object omission (VS in 22):

- (19) *L' ha aperta la ragazza*  
obj.CL AUX open.PP the girl
- (20) *La finestra ha aperto la ragazza*  
the window AUX open.PP the girl
- (21) *Ha aperto la finestra la ragazza*  
AUX open.PP the window the girl
- (22) *Ha aperto la ragazza*  
AUX open.PP the girl

As shown in **Table 6**, speakers from both experimental groups mainly produce answers of the cVS type, thus realizing the object in the appropriate form of a clitic pronoun in VS answers.

The following word orders were found in SV answers: Subject–Verb–Object (SVO in 23), Subject–Clitic–Verb (ScIV in 24), and Subject–Verb answers with object omission (SV in 25):

- (23) *La ragazza ha aperto la finestra*  
the girl AUX open.PP the window
- (24) *La ragazza l' ha aperta*  
the girl obj.CL AUX open.PP

<sup>14</sup>The reasons for excluding answers of the OTHER kind from the analysis is rather straightforward: as reported above, this category mainly includes passives and reduced cleft. In passive structures, the internal argument is promoted to the subject position and is most likely produced as a fully fledged lexical DP, while in reduced cleft of the kind produced by our participants (see example 16 above) the object is just not there.

**TABLE 6 |** Word orders in VS answers (distribution in percentages).

	cVS	OVS	VOS	Omission
AL2S	95.5	0.0	3.2	1.3
HS	87.7	9.6	2.7	0.0

cVS, Clitic–Verb–Subject; OVS, Object–Verb–Subject; VOS, Verb–Object–Subject; omission, no object production.

- (25) *La ragazza ha aperto*  
the girl AUX open.PP

**Table 7** reports which word orders are attested in the SV answers of AL2S and of HS. When SV is the adopted strategy, the object is mainly produced as a full-fledged lexical DP in the postverbal position by both groups. We will comment on this alternative strategy in the discussion section.

Summing up, we can conclude that there is a strong relation between the object form and the adopted answering strategy. Objects are consistently produced as clitic pronouns in VS answers and as lexical DPs in SV answers.

## DISCUSSION

The first result in the collected data is that AL2S and HS do not perform as MonoL1 in their production of NISs. Whereas the latter are very consistent in producing VS structures across all verb types, thus setting a clear benchmark for Italian, multilingual speakers typically access a wider range of options. All AL2S and HS participants produce VS answers, although in different amounts, which overall do not reach the rate of MonoL1 speakers. Among the attested alternative options, the most frequent output is the SV one, with a focalized subject in the preverbal position, namely the one described as prototypical in German answers. Although the distribution of VS and SV answers varies across the two groups also in relation to the kind of verb in use, we can clearly see that the two strategies, i.e., postverbal subject and prosodic prominence, are competing in multilingual speakers. The SV constituent order is certainly grammatical in Italian (which is an SVO language as witnessed by the word order in discourse neutral sentences) and multilingual speakers seem to overextend its use also to contexts with New Information Focus subjects. We surmise that this overextension takes place under the pressure of German.

Although we do not know which answering strategies multilingual speakers would produce in German in the very same conditions (no German version of the test is available), we can still assume that overextension works only in one direction. We do not expect to find VS answers of the Italian kind in their German, because the structure would be simply ungrammatical in this language<sup>15</sup>. In contrast, the possibility to use the German SV answers with NISs (respectively, 26.03% for AL2S, and 44.75% for HS on average) is left open (and probably even favored) by the

<sup>15</sup>In German expletive *pro* and low lexical subjects are limited to impersonal passive structures (Hubert, 1989), which are not relevant in the pragmatic conditions elicited in the present study.

**TABLE 7 |** Word orders in SV answers (distribution in percentages).

	SVO	ScIV	Omission
AL2S	89.3	7.8	2.9
HS	85.3	13.1	1.6

SVO, Subject–Verb–Object; ScIV, Subject–Clitic–Verb; omission, no object production.

grammaticality of the Subject–Verb word order in Italian. Based on these results, we must conclude that the direction of influence is independent of the status of the language (e.g., German as L2 for AL2S or as the majority language for HS), and it rather depends on the characteristics of the languages involved.

We can now answer the first experimental question regarding the comparison among the three groups and conclude that despite the fact that they are all native speakers of Italian, this factor is not sufficient in assuring the production of the same answering strategy in the relevant discourse contexts. The very consistent behavior of MonoL1 speakers is not replicated by the two multilingual groups, in which we rather found optionality. As for HS, we can assume that optionality is determined by the presence of both languages in their linguistic environment since childhood; whereas we claim that optionality emerges in AL2S as an effect of their advanced acquisition of German as L2. The phenomenon therefore qualifies as a form of attrition in AL2S, which manifests itself as an altered system of coding discourse information into the sentence syntactic structure. In sum, we conclude that being a native speaker of the target language is not sufficient *per se* in shaping answering strategies; in contrast, being a multilingual speaker of languages characterized by different strategies crucially leads to access to different strategies and optionality. Nevertheless, this optionality respects the grammatical constraints of the target language.

Moreover, the results of the competition between the two main alternative strategies, i.e., VS and SV, seem to depend on the argument structure of the verbs in use, as shown by the comparison between AL2S and HS in the four syntactic conditions.

The first and most straightforward observation is that there is actually no competition between alternative strategies in existential structures of the *c'è/ci sono* kind ('there is/there are'). The consistent use of the structure does not leave open any possibility for the emergence of non-target-like structures in this condition, and the subject is realized in the postverbal position, mainly as an indefinite DP. Since existential structures with a postverbal subject are unproblematic even in intermediate speakers of L2 Italian, [as reported by Belletti and Leonini (2004) for their German- and French-speaking learners], it would have been very surprising if our participants had produced alternative strategies in the corresponding condition; and indeed this was not the case.

The picture is more articulated in the three remaining syntactic conditions, namely those with transitive, unergative, and unaccusative verbs: although AL2S and HS essentially adopt the same answering strategies, their distribution varies in the two groups according to different patterns. In particular, AL2S and HS show a significant difference in their distribution of SV answers with transitive and unergative verbs, while the discrepancy between the two groups decreases with unaccusative verbs.

The asymmetric distribution of SV answers between AL2S and HS shows that there actually is a persisting difference between the two groups (Mann–Whitney tests revealed differences between AL2S and HS in the production of SV answers, in particular with transitive and unergative verbs, see the “Results” section

for details), such that growing-up as multilinguals and having access to both strategies in the input since childhood plays a role in determining a higher activation of the SV strategy in HS (44.75% across conditions) in comparison to AL2S (26.03% across conditions), and a less frequent activation of the VS strategy (overall at 50.52% in HS, against 61.59% in AL2S). As for AL2S, although VS is still their most active strategy across conditions, they do not behave like monolinguals anymore because multilingualism has enlarged their range of answering strategies.

With respect to the relation between answering strategy and verb type, the pattern becomes clearer when we look at the overall distribution of VS structures: as for AL2S speakers, we see that VS answers are always the preferred strategy from the quantitative point of view, across all verb types; whereas for HS, the number of VS answers increases across the different verb categories. In HS VS answers are at their lowest on transitive verbs (36.82%), they increase with unergative verbs (49.77%), and they are even higher with unaccusative verbs (60.23%); this pattern allows for at least two observations. First, when comparing AL2S and HS the reduced difference between the two groups in sentences with unaccusative verbs is due to a specific increase of HSs' postverbal subjects with respect to the data attested in the other verb categories. Second, both groups produce the lowest number of VS answers with transitive verbs, thus signaling that the presence of a second argument in the structures might have an impact on the subject position too. We analyze both issues in turns.

As for the former, at the core of the unaccusative hypothesis (Perlmutter, 1978; Burzio, 1986; Belletti, 1988; see also Belletti and Bianchi, 2016 for a more recent overview) is the fact that subjects of unaccusative verbs present different properties than the subjects of transitive and unergative verbs; specifically, they are first merged as the internal argument rather than as the external one; moreover, (indefinite) subjects of unaccusatives can be licensed internally to the verb phrase, thus remaining in the postverbal position. Based on the collected data, we assume that the specific property of the verb argument structure favors the production of postverbal subjects in HS. In other words, we suggest that the production of VS structures by HS might be determined, at least in part, by the property of unaccusative subjects *per se*<sup>16</sup>.

As for the second issue concerning transitive verbs, the presence of a second argument in their structure seems to increase optionality; this leads us to the discussion of the fourth research question we raised, namely the one concerning the realization of the object in sentences containing NISs. Under the discourse conditions of the experimental task, the object of transitive verbs is present in the questions together with the verb, thus qualifying as a given Topic in the answer. In the elicited pragmatic condition, Topic-like objects are usually realized as clitic pronouns in Italian. For instance, Leonini and Belletti (2004), elicited answers to questions characterized by topic-like objects from monolingual Italian speakers and observed that

<sup>16</sup>Given the indefiniteness requirement holding on the internal argument of unaccusatives, VS answers with an indefinite subject could have the subject in the VP-internal argument position of its first merge.

participants were very consistent in producing clitic objects in their outputs (91%), and only rarely reproduced the complete lexical DP (7.7%) they had been exposed to in the question. However, in the present study participants' answers only partially meet this expectation: both AL2S and HS produce objects either as clitic pronouns or as full-fledged lexical DPs in comparable amounts. What is most striking in the data is that both groups behave very much alike in their object realization because they all show a strict correspondence between object form and subject position: in VS answers, objects are mainly produced in the form of clitics; while in SV structures, they are produced as lexical DPs in the postverbal position, thus resulting in the SVO word order<sup>17</sup>. Among the alternative solutions, we would like to briefly comment on the fact that HS also produce the OVS word order with lexical objects (9.6% of VS answers), thus realizing a pattern that we interpret as possibly reflecting a V2-type structure in their heritage language. The same strategy is not attested in the AL2S group, thus indicating that attrition has not yet determined the onset of the V2 computation in Italian. If we now consider results with transitives alone, it would be hard to tell whether the object form (clitic/lexical DP) determines the subject position (postverbal/preverbal) or the other way around. However, due to the fact that the production of VS is not at ceiling in the unergative and unaccusative conditions, presence of the object cannot anyway be the only cause for the production of preverbal subjects.

In light of the above discussion with respect to the performance of HS, we would like to claim that the performance on unergative verbs should be taken as the benchmark for the production of postverbal subjects in HSs. Transitive structures reduce the number of postverbal subjects, most likely as a reflex of the presence of the object, mostly when it is realized as a lexical DP and not as a clitic, as its discourse status would require. In contrast, the number of VS answers increases with unaccusative verbs because of the possibly wider source for postverbal subjects (as either internal to verb phrase in the merge position as internal argument of the verb, or as a NIS in the specifier of the low New Information Focus position).

Finally, we would like to point out another aspect of the overall performance of HS participants. Although their production of VS answers is relatively limited, the range of answering strategies they produce is the same as in AL2S. Therefore, there is nothing impoverished in their performance. The observation is relevant because it further supports the claim put forth by Kupisch and Rothman (2016; see also Rinke and Flores, 2014; a.o.) against the description of HS as speakers characterized by an incomplete grammar<sup>18</sup>. As far as the production of answering strategies is concerned, nothing seems to be incomplete or incorrect in their outputs, and the asymmetry must therefore be explained as a consequence of their multilingual grammar(s). The claim that HS

speakers' grammar is not incomplete is based on the comparison with AL2S, who certainly achieved a complete maturation of their native language while growing-up as monolinguals. Since we do not talk about incomplete grammar for AL2S, we do not do so for HS either. However, we could have reached this conclusion, i.e., that HS have an incomplete grammar, if we had only compared their performance to that of MonoL1 speakers: this shows the importance of choosing a richer array of speakers to which HS should be compared. Note that AL2S represent the prototypical source of input for HS in the familial environment; under this perspective, the similarity between AL2S and HS can be further evaluated as the sign of a rather successful acquisition of the target language by HS based on the quality of the input they received. The mild differences between AL2Ss and HS rather relate to the multilingual setting of the HS linguistic development.

## CONCLUSION

As to the variables, we considered and which are factors contributing to language dominance, we can conclude that nativeness, multilingualism and amount of input in L1 acquisition do not play equivalent roles when seen through the lenses of answering strategies.

In the present study, we showed that AL2S and HS have access to a wider range of answering strategies with respect to monolingual speakers, thus indicating that nativeness does not guarantee the production of homogeneous answering strategies (as attested in monolingual speakers). This leads to the second variable we considered, i.e., multilingualism: we interpret the presence of SV answers in AL2S as a consequence of the multilingual competence they developed in their adult life; the activation of the German SV word order while computing NISs in Italian shows that alternative strategies can take over the VS strategy in place from childhood (Belletti, 2007), thus resulting in attrited performance in L1. Based on these results, we conclude that language dominance within specific linguistic phenomena can possibly be undermined throughout the lifespan under the pressure of advanced L2 acquisition.

The role of the amount of input in L1 acquisition emerges from the data on the use of alternative strategies: since the two groups differ in the distribution of SV answering strategies, we claim that the condition in which L1 Italian is acquired (as the only L1 for AL2S or as the heritage language for HS) plays a role in determining how active the SV strategy is in Italian, and therefore also how (un)stable the production of postverbal NISs is in multilingual speakers. In particular, differences between AL2S and HS in the production of answering strategies can be better explained under an analysis that takes into consideration the argument structures of the verbs in use, with unaccusative verbs particularly favoring the production of postverbal subjects in HS.

Overall, optionality between VS and SV is attested in all multilingual speakers. Based on these results, we claim that the competence of HS is target-like because AL2S represent their typical source of input during early language development; therefore, nothing is missing, incorrect or incomplete in the grammar of HS.

<sup>17</sup>The VOS order, only marginally possible in Italian, is attested very sporadically, in the HS group only. The VSO order, impossible in Italian for principled reasons (Belletti, 2004), is totally absent in both groups. This is an interesting convergence with the results in the original study on L2 Italian by Belletti and Leonini (2004).

<sup>18</sup>See also Di Venzio et al. (2016) and Schmitz et al. (2016), who argue against incomplete acquisition of the heritage language in Italian-German HS who grew up in Germany. Their claim builds on the lack of significant differences with respect to monolinguals when the production of lexical/pronominal objects and that of null/overt subjects are analyzed in their spontaneous speech.

Finally, the analysis of answers with transitive verbs showed a relation in both groups between the subject position and the object form: AL2S and HS consistently produced object clitic pronouns in combination with postverbal subjects, and lexical object DPs in combination with preverbal subjects. Again, the performance of multilingual speakers differed from what is expected in monolingual speakers (i.e., Topic-like objects realized as clitic pronouns), thus suggesting that this further discourse-related structure is an unstable domain in multilingualism. The phenomenon is well-known from L2 acquisition and should be further explored in future research in the domain of L1 attrition and heritage language, with a particular attention to its interaction with other arguments in the sentence. This is the topic of current research.

## ETHICS STATEMENT

Ethics approval for this research was not required as per the guidelines of the Faculty of Modern Languages at Goethe University as well as per national regulations. The test did not involve any risk on participants' physical, mental and social

integrity. All adult subjects gave written informed consent to participate in the study and could withdraw consent to use the collected data at any time without reprisal.

## AUTHOR CONTRIBUTIONS

AB conceived the idea of the study and wrote the Sections "Answering Strategies" and "Subjects at the Interfaces: Previous Results From Multilingual Speakers." IC collected the data and wrote the Sections "Research Questions," "Materials and Methods," "Results," and "Discussion." CP wrote the Sections "Introduction" and "Conclusion."

## FUNDING

AB's 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."

## REFERENCES

- Belletti, A. (1988). The Case of unaccusatives. *Linguist. Inq.* 19, 1–34.
- Belletti, A. (2001). "Inversion as focalization," in *Subject Inversion in Romance and the Theory of Universal Grammar*, eds A. Hulk and J. Y. Pollock (Oxford: Oxford University Press on Demand), 60–90.
- Belletti, A. (2004). "Aspects of the low IP area," in *The Structure of CP and IP. The Cartography of Syntactic Structures*, Vol. 2, ed. L. Rizzi (Oxford: Oxford University Press on Demand), 16–51.
- Belletti, A. (2005). "Answering with a 'cleft': The role of the null subject parameter and the VP periphery," in *Proceedings of the XXX Incontro di Grammatica Generativa*, (Italy: University of Siena) 63–82.
- Belletti, A. (2007). "Answering strategies. A view from acquisition," in *Romance Languages and Linguistic Theories 2005*, eds S. Baauw, F. Drijkoningen, and M. Pinto (Amsterdam: John Benjamins Publishing Company), 19–38. doi: 10.1075/cilt.291.04bel
- Belletti, A. (2009). *Structures and Strategies*. Abingdon: Routledge.
- Belletti, A. (2015). "The focus map of clefts: extraposition and predication," in *Beyond Functional Sequence*, Vol. 10, ed. U. Shlonsky (Oxford: Oxford University Press), 42–60.
- 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. doi: 10.1007/s11049-007-9026-9
- Belletti, A., and Bianchi, V. (2016). "Definiteness effect and unaccusative subjects: an overview and some new thoughts," in *Definiteness Effects: Bilingual, Typological and Diachronic Variation*, eds S. Fischer, T. Kupisch, and E. Rinke (Cambridge: Cambridge Scholars Publishing), 14–65.
- Belletti, A., and Leonini, C. (2004). "Subject inversion in L2 Italian," in *EUROSLA Yearbook*, Vol. 4, eds S. Foster-Cohen, M. Sharwood Smith, A. Sorace, and O. Mitsuhiro (Amsterdam: John Benjamins), 95–118.
- Belletti, A., and Rizzi, L. (2017). "On the syntax and pragmatics of some clause-peripheral positions," in *Formal Models in the Study of Language*, eds J. Blochowiak, C. Grisot, S. Durrleman, and C. Laenzlinger (Berlin: Springer), 52–75.
- Benincà, P., and Poletto, C. (2004). "Topic, focus, and V2," in *The Structure of CP and IP. The Cartography of Syntactic Structures*, Vol. 2, ed. L. Rizzi (Oxford: Oxford University Press on Demand), 52–75.
- Burzio, L. (1986). *Italian Syntax: A Government-Binding Approach*, Vol. 1, Dordrecht: Springer. doi: 10.1007/978-94-009-4522-7
- Cardinaletti, A. (2004). "Towards a cartography of subject positions," in *The Structure of CP and IP. The Cartography of Syntactic Structures*, Vol. 2, ed. L. Rizzi (Oxford: Oxford University Press on Demand), 115–165.
- Cardinaletti, A., and Starke, M. (1999). "The typology of structural deficiency: a case study of the three classes of pronouns," in *Clitics in the Languages of Europe*, ed. H. van Riemsdijk (Berlin: Mouton de Gruyter), 145–233.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Cruschina, S. (2009). *Discourse-Related Features and the Syntax of Peripheral Positions: A Comparative Study of Sicilian and Other Romance Languages*. Ph.D. thesis, University of Cambridge, Cambridge.
- Cruschina, S. (2012). *Discourse-Related Features and Functional Projections*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199759613.001.0001
- Dal Pozzo, L., and Guesser, S. (2010). A comparative view on answering strategies and new information subjects in Brazilian Portuguese and Finnish. *Stud. Linguist.* 4, 82–107.
- Di Venanzio, L., Schmitz, K., and Scherger, A. L. (2016). Objects of transitive verbs in Italian as a heritage language in contact with German. *Linguist. Approaches Biling.* 6, 227–261. doi: 10.1075/lab.13041.div
- Gabriel, C. (2010). On focus, prosody, and word order in Argentinean Spanish: a Minimalist OT account. *Rev. Virtual Estud. Ling.* 4, 183–222.
- Genevska-Hanke, D. (2017). *Subject Realization in Bulgarian. Overt and Null Subjects in Bulgarian German Interlanguage*. Ph.D. thesis, University of Oldenburg, Oldenburg.
- Grüter, T., and Paradis, J. (eds) (2014). *Input and Experience in Bilingual Development*, Vol. 13, Amsterdam: John Benjamins Publishing Company.
- Hamann, C., and Tuller, L. (2014). Genuine versus superficial relatives in French: the depth of embedding factor. *Rev. Grammatica Generativa* 36, 146–181.
- Hoot, B. (2012). *Presentational Focus in Heritage and Monolingual Spanish*. Ph.D. thesis, University of Illinois at Chicago, Chicago, IL.
- Hubert, H. (1989). "Theta-tracking systems. Evidence from German," in *Configurationality*, eds L. Maráz and P. Muysken (Dordrecht: Foris Publication), 185–206.
- Jaeggli, M., and Safir, K. (eds.) (1989). *The Null Subject Parameter*, Vol. 15, Berlin: Springer Science & Business Media.
- Kupisch, T., and Rothman, J. (2016). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Biling.* 5, 564–582. doi: 10.1177/1367006916654355
- Kupisch, T., and Weijer, J. V. D. (2016). "The role of the childhood environment for language dominance: a study of adult simultaneous bilingual speakers



- of German and French,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Dallers (Cambridge: Cambridge University Press), 174–194.
- Labuz, K. (2012). *Soggetto Di Nuova Informazione: Un Confronto Sperimentale Tra Polacco e Italiano*. Ph.D. thesis, University of Siena, Siena.
- Leal, T., Destruel, E., and Hoot, B. (2017). The realization of information focus in monolingual and bilingual native Spanish. *Linguist. Approaches Biling.* 8, 217–251. doi: 10.1177/0267658318784343
- Leonini, C., and Belletti, A. (2004). “Adult L2 acquisition of Italian clitic pronouns and ‘subject inversion’/VS structures,” in *Proceedings of Gala 2003*, eds J. van Kampen and S. Baauw (Utrecht: LOT Occasional Series 3), 293–304.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk: Transcription Format and Programs*, 3rd Edn. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Montrul, S. (2016). “Dominance and proficiency in early and late bilingualism,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Dallers (Cambridge: Cambridge University Press), 15–35.
- O’Grady, W., Lee, O. S., and Lee, J. H. (2011). Practical and theoretical issues in the study of heritage language acquisition. *Herit. Lang. J.* 8, 23–40.
- Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. *Annu. Meet. Berkeley Linguist. Soc.* 4, 157–190. doi: 10.3765/bls.v4i0.2198
- 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
- Rinke, E., and Flores, C. (2014). Morphosyntactic knowledge of clitics by Portuguese heritage bilinguals. *Biling. Lang. Cogn.* 17, 681–699. doi: 10.1017/S136672891300076X
- Rinke, E., and Flores, C. (2018). Another look at the interpretation of overt and null pronominal subjects in bilingual language acquisition. Heritage Portuguese in contact with German and Spanish. *Glossa* 3:68. doi: 10.5334/gjgl.535
- Rizzi, L. (1982). *Issues in Italian Syntax*, Vol. 11, Berlin: Walter de Gruyter. doi: 10.1515/9783110883718
- Rizzi, L. (1997). “The Fine Structure of the Left Periphery,” in *Elements of Grammar*. Dordrecht: Springer, 281–337. doi: 10.1007/978-94-011-5420-8\_7
- Rizzi, L., and Shlonsky, U. (2007). “Strategies of subject extraction,” in *Interfaces + Recursion = Language?*, eds U. Sauerland and H. M. Gärtner (Berlin: De Gruyter Mouton), 115–160.
- 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
- Schmitz, K., Di Venzio, L., and Scherger, A. L. (2016). Null and overt subjects in Italian and Spanish heritage speakers in Germany. *Lingua* 180, 101–123. doi: 10.1016/j.lingua.2016.04.004
- Sorace, A. (2005). “Selective optionality in language development,” in *Syntax and Variation: Reconciling the Biological and the Social*, eds L. Cornips and K. P. Corrigan (Amsterdam: John Benjamins Publishing), 55–80. doi: 10.1075/cilt.265.04sor
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguist. Approaches Biling.* 1, 1–33. doi: 10.1075/lab.1.1.01sor
- Sorace, A., Serratrice, L., Filiaci, F., and Baldo, M. (2009). Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children. *Lingua* 119, 460–477. doi: 10.1016/j.lingua.2008.09.008
- Tsimpli, I., Sorace, A., Heycock, C., and Filiaci, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *Int. J. Biling.* 8, 257–277. doi: 10.1177/13670069040080030601

**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 Caloi, Belletti and Poletto. 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.



# Age of Onset and Dominance in the Choice of Subject Anaphoric Devices: Comparing Natives and Near-Natives of Two Null-Subject Languages

Elisa Di Domenico<sup>1\*</sup> and Ioli Baroncini<sup>1,2</sup>

<sup>1</sup> Dipartimento di Scienze Umane e Sociali, Università per Stranieri di Perugia, Perugia, Italy, <sup>2</sup> Scuola Superiore di Dottorato e di Specializzazione, Università per Stranieri di Siena, Siena, Italy

## OPEN ACCESS

### Edited by:

Dobrinka Genevska-Hanke,  
University of Oldenburg, Germany

### Reviewed by:

Lena Dal Pozzo,  
Università degli Studi di Firenze, Italy  
Ianthi Tsimpli,  
University of Cambridge,  
United Kingdom

### \*Correspondence:

Elisa Di Domenico  
elisa.didomenico@unistrapg.it

### Specialty section:

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

**Received:** 27 March 2018

**Accepted:** 18 December 2018

**Published:** 10 January 2019

### Citation:

Di Domenico E and Baroncini I  
(2019) Age of Onset and Dominance  
in the Choice of Subject Anaphoric  
Devices: Comparing Natives  
and Near-Natives of Two Null-Subject  
Languages. *Front. Psychol.* 9:2729.  
doi: 10.3389/fpsyg.2018.02729

Several studies have highlighted the role of cross-linguistic influence in determining the over-use of overt subject pronouns in near-native speakers of a null-subject language as Italian. In this work we inquire on the role of factors different from cross-linguistic influence in the choice of anaphoric devices in near-natives, such as age of onset of exposure and dominance. In order to do so, comparing the productions of two groups of natives speakers, we first single out two null-subject languages, Italian and Greek, which do not differ significantly as far as subject anaphoric devices are concerned and thus instantiate a suitable language combination to investigate the role of factors other than cross-linguistic influence in bilingual speakers of these two languages (Study 1). In Study 2, we compare the productions of a group of native speakers and two groups of near-native speakers in Italian: Greek-Italian bilinguals from birth and L2ers of Italian with Greek as an L1. Results reveal that over-use of overt pronouns in near-natives occurs in the absence of cross-linguistic influence and that age of onset of exposure is a relevant factor: while bilinguals from birth do not differ from native speakers, L2ers over-use overt pronouns compared to both native speakers and bilinguals from birth. In order to establish whether dominance is a possible factor determining bilinguals' choice of subject anaphoric devices, in Study 3, we compare two groups of Greek-Italian bilinguals from birth: bilinguals living in Greece (whose predominant language is Greek) and bilinguals living in Italy (whose predominant language is Italian). Results reveal no effect of dominance in the production of overt subject pronouns. We found, however, an unexpected effect in the predominant language of one group: bilinguals living in Greece produce significantly more null pronouns and less lexical DPs in Greek compared to bilinguals living in Italy. We interpret this effect as stemming from the need to differentiate the two languages that these bilingual speakers have to handle in everyday life. Interestingly, this effect is found in the predominant language rather than in the non-predominant one.

**Keywords:** age of onset, dominance, Italian, Greek, overt subject pronouns, null subject pronouns (*pro*), natives, near-natives

## INTRODUCTION

Some languages of the world are null-subject languages. In these languages the subject of finite clauses (whether matrix or embedded) can be left unpronounced, as in (1.b/d), (2.b/d) and (3.b/d):

- |  |                    |
|--|--------------------|
| (1) a. Gianni ha parlato<br>G. spoke<br>b. <i>pro</i> Ha parlato<br>He spoke<br>c. Lui ha parlato<br>He spoke<br>d. Gianni ha detto che <i>pro</i> ha parlato<br>G. said that he spoke | Italian            |
| (2) a. Juan habló<br>b. <i>pro</i> habló<br>c. Él habló<br>d. Juan dijo que <i>pro</i> habló   | Spanish            |
| (3) a. O Janis milise/ Milise o Janis<br>b. <i>pro</i> Milise<br>c. Aftos milise<br>d. O Janis ipe oti <i>pro</i> milise   | Greek <sup>1</sup> |

Though phonetically unrealized, the null subject is syntactically active, and is standardly indicated as *pro*, as shown in the .b and .d examples above.<sup>2</sup> Given that null-subject languages have both overt (as shown in the c. examples above) and null subject pronouns, an interesting question is what the division of labor is between the two series of pronouns.

Calabrese (1986) for instance has noted that in Italian, in cases like (4), the null pronoun takes the antecedent in subject position, while the overt pronoun preferentially takes an antecedent which is not the subject:

- (4) a. Quando Carlo<sub>i</sub> ha picchiato Antonio<sub>j</sub> *pro*<sub>i/\*j</sub> era ubriaco  
b. Quando Carlo<sub>i</sub> ha picchiato Antonio<sub>j</sub> lui<sub>j/\*i</sub> era ubriaco  
When C. hit A *pro*/he was drunk

<sup>1</sup>Greek third person personal pronouns disappeared around the 5th–6th century B.C. (Panagiotidis, 2000) and were substituted by demonstratives, as *aftos* in (3.c; this one) or *ekinos* (that one), with an anaphoric function. These demonstrative pronouns can also have an inanimate antecedent, contrary to Italian overt pronouns. In Italian, demonstrative pronouns can also be used with anaphoric function, with inanimate referents only or, with a pejorative flavor, in sub-standard varieties with animate/human referents. (3.a) shows another difference between Greek and Italian, in that Greek freely allows post-verbal subjects (Roussou and Tsimpli, 2006), and VSO, while in Italian the post-verbal position of subjects is restricted to new-information focus subjects and VSO is impossible (Belletti, 2001, 2004). Post-verbal subjects and VSO are also possible in Spanish, although Greek and Spanish partly differ, word order being more flexible in Greek than in Spanish, since VSO is allowed by different mechanisms in the two languages (Roussou and Tsimpli, 2006).

<sup>2</sup>The null pronoun, for instance, binds the anaphor *se stesso* in (i.c) as the lexical DP *Gianni* does in (i.a). Contrary to pronouns (*lo* in i.b), anaphors must be bound within the clause containing them:

- (i) a. Paolo<sub>i</sub> ha detto [che [Gianni<sub>j</sub> proteggerà se stesso<sub>j</sub>]]  
P. said that G. will protect himself  
b. Paolo<sub>i</sub> ha detto [che [Gianni<sub>j</sub> lo-<sub>j</sub> proteggerà]]  
P. said that G. him will protect  
c. Paolo<sub>i</sub> ha detto [che [*pro*<sub>i</sub> proteggerà se stesso<sub>i</sub>]]  
P. said that he will protect himself

Noting that a post-verbal subject cannot be the antecedent of a pronoun (whether null or overt, as shown in (5)) and that *pro* can co-refer with the dative PP of so called *Psych*-verbs in preverbal position (6), the author proposes that the property ‘subject’ is not sufficient to characterize the referential properties of *pro*:

- (5) a. \*Ha parlato Carlo<sub>i</sub> quando *pro*<sub>i</sub> è arrivato  
b. \*Ha parlato Carlo<sub>i</sub> quando lui<sub>i</sub> è arrivato.  
Spoke C. when *pro*/he arrived  
(6) Poiché a Giovanni<sub>i</sub> piace Maria, *pro*<sub>i</sub> fa di tutto per farsi bello ai suoi occhi  
Because G. likes M. he does everything to show off for her

Calabrese (1986) proposes that the relevant property is instead ‘Subject of primary predication’ (or Thema).<sup>3</sup>

As far as overt pronouns (‘stressed’ in his terms) are concerned, Calabrese (1986) assumes that they are only used when the occurrence of their referent is not expected, proposing a principle like (7):

- (7) Assign the feature [+ stressed] to a pronominal X only when the occurrence of the referent of X is not expected [Calabrese, 1986: 7, ex. (18)]

Assuming that expectedness (i.e., high probability of occurrence) is correlated to low content of information, while unexpectedness (i.e., low probability of occurrence) is correlated to high content of information, he argues that (7) simply prevents giving more information than is required, and hence is a direct consequence of the second maxim of quantity of Grice (1975).<sup>4 5</sup>

We may thus easily derive from (7) the fact that overt pronouns, at least in Italian and Greek, are required only in case of topic shift or focalization, i.e., when their referent is unexpected. But when the referent is expected, overt pronouns are impossible:

- (8) a. Poiché *pro*<sub>i</sub> ha visto quel film, Mario<sub>i</sub> si è spaventato  
b. \*Poiché lui<sub>i</sub> ha visto quel film, Mario<sub>i</sub> si è spaventato  
Because *pro*/he saw that film, M. was frightened  
[Calabrese, 1986 ex. (19) and (23)]  
(9) a. Epidi *pro*<sub>i</sub> ide ekini tin tenia, o Marios<sub>i</sub> tromaxe.  
b. \*Epidi aftos<sub>i</sub> ide ekini tin tenia, o Marios<sub>i</sub> tromaxe. <sup>6</sup>

Things appear to work in part differently for near-native speakers, as brought to light by a number of studies. While a natural reply to (10.A) would be (10.B1) for a native speaker, near-natives may also produce (10.B2):

<sup>3</sup>Experimental findings by Carminati (2002) suggest indeed that *pro*, at least in intra-sentential anaphora, looks for an antecedent in Spec, IP. Subjects of predication share properties with topics: according to Rizzi (2005, 2018), subjects and topics share an ‘aboutness’ property. According to Lambrecht (1994: 118) topics are ‘the thing which the proposition expressed by the sentence is about.’

<sup>4</sup>Do not make your contribution more informative than is required’ [Grice, 1975: 45]. Calabrese (1986: fn. 6) suggests that the Avoid Pronoun Principle (Chomsky, 1981: 65) must be interpreted in a similar vein.

<sup>5</sup>A similar claim is made by Chiou (2013) for Greek.

<sup>6</sup>Note that (8.b) and (9.b) are possible with a different indexing, i.e., if *lui/aftos* does not co-refer with *Mario*.

- (10) A. Perché Giorgio si è licenziato?  
 Why did G. resign  
 B1. Perché *pro* non sopportava più il direttore  
 B2. Perché lui non sopportava più il direttore  
 Because *pro*/he could not stand the boss anymore  
 [Adapted from Sorace, 2006: 507]

Tsimpli et al. (2004) for instance studied the production and comprehension of overt and null subject pronouns by native speakers of Italian and native speakers of Greek who were near-native speakers of English as an L2 and had a minimum of 6 years of residence in Britain. They were hence experiencing attrition from the L2.<sup>7</sup> As for the Italian experimental subjects, the authors found a significant difference between the control and the experimental group in the choice of the matrix subject as a possible referent of the overt pronoun in the embedded sentence.<sup>8</sup>

Sorace and Filiaci (2006) studied the comprehension of null and overt subject pronouns in Italian by English speakers who had learned Italian as adults, reaching a near-native level of proficiency. Compared to native speakers, near-natives had a significantly higher preference for the subject of the matrix clause as a possible antecedent of overt subject pronouns.<sup>9</sup>

Belletti et al. (2007) were also concerned with near-native speakers of Italian whose native language was English, and who had started learning Italian as adults. Their findings on pronoun comprehension and production matched: overt pronouns were over-produced and also interpreted in co-reference with a topical antecedent by these near-native speakers.

Serratrice et al. (2004) studied the productions of an Italian-English bilingual child, finding an overuse of overt pronouns in her Italian.<sup>10</sup>

Taken together these studies support the idea that the over-use and over-interpretation of overt pronouns is due to cross-linguistic influence from English, a language which has only overt pronouns. But then the question is why the influence goes only from English to Italian and not in the other direction. One possibility is that these speakers chose the option compatible with

both their languages: coherently with Hulk and Müller's (2000) hypothesis, cross-linguistic influence does not occur in young bilinguals unless input from one of the languages can be analyzed through the grammar of the other language. Another possibility is, however, that overt pronouns are, for some reason, 'simpler' for speakers of more than one language: if so, they should be over-produced (and over-interpreted) also in the absence of cross-linguistic influence.

Sorace and Filiaci (2006: 345) quote production data collected by Bini (1993) from low-intermediate Spanish learners of Italian who use overt pronouns in contexts in which both Italian and Spanish would require a null pronoun: since cross-linguistic influence cannot be implicated in this case, the authors suggest that overt pronouns may be a default form.

Sorace et al. (2009) compare the preferences toward null and overt subject pronouns in Italian, in a [+Topic Shift] and [-Topic Shift] condition by different groups of subjects: Italian monolingual adults, Italian monolingual children, English-Italian bilingual children (6–7 and 8–10 years old, living in Italy and living in the United Kingdom), Spanish-Italian bilingual children.

In the [-TS] condition younger children chose significantly more overt pronouns than older children and adults, and older children more than adults. Children with English as the community language were more likely to choose inappropriate overt pronouns than children with Italian as the community language at the age of 6–7, but not at 8–10. Italian monolingual children aged 6–7 chose significantly more overt pronouns than adults. Spanish-Italian bilinguals were significantly more likely to opt for an overt pronoun than the monolinguals, but they were not significantly different from the English-Italian bilinguals. In the [+TS] condition bilingual children (regardless of the language combination) accepted more null subject pronouns than monolingual children.<sup>11</sup>

These results are very important in that they show that establishing the appropriate conditions for pronoun resolution is a phenomenon which is acquired late, in part independently from cross-linguistic influence in bilingual children, since Spanish-Italian bilingual children behaved differently from Italian monolingual children. These results also show that the pattern is not completely asymmetric, given some variability in the acceptance of null pronouns in [+TS] contexts.

The fact that the preferences of Spanish-Italian bilingual children may not be due to cross-linguistic influence has been challenged, however, by a self-paced reading study on Spanish and Italian (Filiaci et al., 2013) that found that pronominal preferences may not be the same in Italian and Spanish, although they are both null-subject languages. Sentences containing an overt pronoun congruent with a complement antecedent (as in (11)) were read significantly faster in Italian, but not in Spanish, suggesting that overt pronouns in Spanish are also compatible with a topic antecedent:

<sup>11</sup>The authors propose difficulties at integrating different types of information in real time as an explanation for their results. Along the same line, Sorace (2011) suggests that there could also be a difference in the processing resources available for bilingual and monolingual speakers.

<sup>7</sup>The study was also concerned with post-verbal subjects, which are possible in null-subject languages but not in non-null subject languages, as originally noted by Rizzi (1982). As noted in footnote 1, Greek and Italian, however, differ in this respect.

<sup>8</sup>Results concerning the interpretation of null and overt pronouns are presented in Tsimpli et al. (2004) only for the Italian participants. Experimental sentences were of the kind given in (i.a) and (i.b):

- (i) a. L'anziana signora saluta la ragazza quando lei attraversa la strada  
 The old woman greets the girl when she is crossing the street  
 b. Quando lei attraversa la strada, l'anziana signora saluta la ragazza  
 When she is crossing the street, the old woman greets the girl

Results were particularly clear in the (i.a) condition. These results concerning the comprehension of overt pronouns are however not matched in the production tasks, such as the Story Telling task, for which, as the authors acknowledge, no significant results were attested for either group (Tsimpli et al., 2004: 267).

<sup>9</sup>Experimental materials were very similar to those employed by Tsimpli et al. (2004). Here as well, results were particularly clear in cases like (i.a) of footnote 8.

<sup>10</sup>Several studies tackle indeed this issue examining spontaneous productions of bilingual children with a null and a non-null subject language. See, among others, Paradis and Navarro (2003) on a Spanish-English child, Pinto (2006) on two Dutch-Italian children, Hacoen and Schaeffer (2007) on a Hebrew-English child. They all found an over-use of overt subjects in the null-subject language.



- (11) a. Dopo che Giovanni<sub>i</sub> ha criticato Bruno<sub>j</sub> così ingiustamente, lui<sub>i</sub> si è sentito offeso  
 b. Despues de que Bernardo<sub>i</sub> criticó a Carlos<sub>j</sub> tan injustamente, él<sub>i</sub> se sintió muy ofendido.  
 After that G./B.<sub>i</sub> has criticized B./C.<sub>j</sub> so unjustly, he<sub>i</sub> felt offended

This makes the authors explicitly claim that the findings in Sorace et al. (2009) concerning the preference differences of Spanish-Italian bilingual children compared to Italian monolingual children could indeed be due to cross-linguistic influence from Spanish (Filiaci et al., 2013: 17).

This suggests that in order to verify whether the over-use/ over-acceptance of overt subject pronouns in bilinguals is not only due to cross-linguistic influence, care must be put in the choice of the language combination of bilingual speakers, since not all null-subject languages are alike in this respect.

In this work we present three studies concerning adult narrative productions in Italian and Greek by two groups of native speakers (Italian natives and Greek natives), two groups of adult Italian-Greek bilinguals from birth (Bilinguals living in Greece and Bilinguals living in Italy) and a group of adult native speakers of Greek who started to learn Italian in adulthood reaching a near-nativeness level of proficiency (L2ers).

In Study 1, we compare the productions of the two groups of native speakers, highlighting that there are no significant quantitative differences in Greek and Italian as far as the implementation of null pronouns, overt pronouns and lexical DPs are concerned, so that Italian and Greek appear as a suitable language combination to study the factors influencing bilinguals' choices of anaphoric devices, in the absence of effects related to cross-linguistic.

In Study 2, we compare the productions in Italian of a group of native speakers and two groups of near-native speakers: bilinguals from birth and L2ers. Results reveal that near-natives over-use overt pronouns also when cross-linguistic influence is absent and that age of onset of exposure to Italian is a relevant factor in this respect: while bilinguals from birth do not differ from native speakers, L2ers over-use overt pronouns compared to both native speakers and bilinguals from birth.

In order to establish whether dominance is a possible factor determining speakers' choice of anaphoric devices, in Study 3, we compare two groups of bilinguals: bilinguals living in Greece and bilinguals living in Italy. Results reveal no effect of dominance with respect to the production of overt-pronouns, neither in Italian nor in Greek. We found, however, an unexpected effect in the predominant language of one of the groups: bilinguals living in Greece produce significantly more null pronouns and less lexical DPs in Greek compared to bilinguals living in Italy. We interpret this effect as stemming from the need to differentiate the two languages that this bilingual group has to handle in everyday life. Interestingly, this effect is found in the predominant language rather than in the non-predominant one, and does not concern overt pronouns.

## STUDY 1: SUBJECT ANAPHORIC DEVICES IN ITALIAN NATIVES AND GREEK NATIVES

The study conducted by Filiaci et al. (2013) reviewed in the previous section suggests that an analogous null/overt pronouns division of labor among null-subject languages should not be taken for granted. Spanish, as the authors show, differs from Italian in that overt pronouns appear to retrieve a subject antecedent to a greater extent in Spanish compared to Italian. In Study 1, we therefore compare the productions of two groups of native speakers (Italian native speakers and Greek native speakers) in order to see whether the proportion of null and overt pronouns and lexical DPs produced is comparable in the two groups. If this analysis reveals no significant differences, differences in the productions of speakers of the two languages could not be attributed to cross-linguistic influence.

### Subjects

20 subjects participated in Study 1: 10 native speakers of Italian and 10 native speakers of Greek.

Italian Natives (6 male; 4 female) had a mean age of 32 (range 19–58). They were born in Italy and had been living there by the end of testing. Three of them had a university degree, while seven had a high school degree and were attending university.

Greek Natives (4 male; 6 female) had a mean age of 29 (range 19–58). They were born in Greece and had been living there by the end of testing. Four of them had a university degree, while six had a high school degree and were attending university.

## Materials and Methods

### Ethical Considerations

There is no ethical committee in our institutions, and for this reason this study could not undergo an ethical reviewing process, not required according to the guidelines of our institution and national regulations in such cases. The subjects in this study were adults who participated in it on a voluntary basis and came to the place of data collection for this purpose only. They were informed about the general aims of the research and gave their written informed consent to the treatment of the data they produced, including the publication of the results. In order to protect their anonymity, subjects were coded only by progressive numbers in the data analysis.

### Procedure

Subjects were asked to watch a short movie (The Pear Film) and then tell the story.<sup>12</sup> Subjects productions were recorded and then transcribed with the help of the CLAN system (part of the CHILDES tools, MacWhinney, 2000). Subjects were tested individually in a quiet room and the interviewer did not interact with them during their narration.

<sup>12</sup>The Pear Film is a 6-min film without dialogs. It was created at the University of California at Berkeley in 1975 by a group of linguists to collect narration data. See Chafe (1980) for a first report of this research. Years later, data collected through the Pear Film have become part of the experimental material in works dedicated to the study of pronoun production and resolution in bilingual contexts, such as Tsimpli et al. (2004) and Belletti et al. (2007).

## Defining the Reference Total

The narrations collected with the procedure described above were then analyzed in order to study the occurrences of null and overt subject pronouns as well as of subject lexical DPs chosen by the speakers. Given the nature of the task (semi-spontaneous production), the two corpora contained a great variety of clausal types. Not all of them, however, can be considered suitable environments to study speakers' choice of subject referring expression, since in many of these clausal types no true clause-internal choice is possible as far as their subject is concerned, since it is syntactically determined. For instance, this is the case in subject relatives, where, according to a raising analysis of this clausal type, the subject is the copy of the moved head of the relative, or in pseudo-relatives, where the antecedent must be overt and the internal subject is invariably null. In subject clefts the subject is focalized, hence it cannot be null. As for absolute gerundive and participial, adjectival and prepositional small clauses, their subject is standardly assumed to be PRO. The subject is also syntactically determined in Italian infinitives (whether control, raising or ACC-ing) and in Greek *na* and *ke* clauses, when they are complement of certain verbs.<sup>13</sup> Finally, the subject of existential sentences is syntactically determined, in Italian as well as in Greek.<sup>14</sup>

For this reason, we kept in what we call the 'Reference Total' only those clausal types whose subject can be chosen clause-internally by the speaker, i.e., finite and copular sentences as well as non-subject relatives and non-subject clefts.

Since we adopted this 'free clause-internal choice' criterion, other cases had to be excluded, as well.

Finite sentences whose subject was the narrator or included narrator+ interviewer were also excluded, since they were in the first person (singular or plural), and a choice between a null pronoun, an overt pronoun or a lexical DP is only possible in the third person, lexical DPs being excluded from first and second person.<sup>15</sup>

<sup>13</sup>Greek doesn't have infinitives, but rather embedded sentences introduced by *na* or *ke* complementizers whose verbs are inflected. Verbs in the matrix clause that embed a *na* or *ke* complement clause are perception, knowledge aspectual and modal verbs (for a complete list see e.g., Ingria, 2005; Spyropoulos, 2007). For these cases, there is disagreement in the relevant literature as to the kind of subjects inside these clausal types (see e.g., Philippaki-Warbuton and Catsimali, 1999; Alexiadou and Anagnostopoulou, 2002; Spyropoulos, 2007 among others, and the references quoted there) but all analyses agree on the fact that inside these complement clauses the subject (as well as tense) is dependent on the subject of the matrix clause and can never be overt. *na* clauses can also occur as independent clauses and in this case they are considered subjunctive clauses (hence with an independent subject, which can be null or overt). Matrix clauses with verbs like *elpizo* (hope), *perimeno* (wait/expect), *pistevo* (believe) embed *na* clauses whose subject (and tense) is independent from the one of the matrix clause (Spyropoulos, 2007).

<sup>14</sup>Greek has two kinds of existential sentences, those involving the verb *ine* (be-3sg/pl) and those involving the verb *echi* (have-3sg), whose subject can never be overt. Italian existential sentences contain the locative clitic *ci* and the verb *essere* (be), and a so called 'pivot' which can be definite or indefinite. In the type of existential most attested in our corpus, the one containing an indefinite pivot, the (expletive) subject is assumed to be *ci*.

<sup>15</sup>Some of the sentences with a 'narrator' subject were indeed stock phrases such as Gr. *xero go* (I don't know, lit. know I) or It. *diciamo* (let's say).

Some of the sentences were used to introduce (rather than to resume) a Discourse Referent, and since first mention is always lexical, we excluded those sentences as well.

In this way we obtained the Reference Total, which consists of 387 sentences produced by the Italian natives and 454 sentences produced by the Greek natives. In this Reference Total we analyzed the occurrences of null pronouns, overt pronouns and lexical DPs.

## Results<sup>16</sup>

Null subject pronouns are the most employed anaphoric device (67.18% by Italian natives; 69.38% by Greek natives), followed by lexical DPs (24.28% by Italian natives; 23.12% by Greek natives), while overt pronouns are quite rare (6.20% in the Italian natives Reference Total; 3.37% in the Greek natives Reference Total). We have singled out another resumption device which we call 'other' and which consists of various quantificational expressions such as It. '*uno*' (one), '*uno dei tre*' (lit. one out of the three), '*tutti*' (all of them), Gr. '*enas apo aftous*' (one of them). Instances of 'other' are quite rare, as well (2.06 % in the Reference Total of the Italian natives; 3.74% in the Reference Total of the Greek natives).

A  $\chi^2$ -test reveals no significant difference between the two groups neither for *pro* ( $\chi^2 = 0.4675$ , n.s.) nor for lexical DPs ( $\chi^2 = 0.1561$ , n.s.), overt pronouns ( $\chi^2 = 2.2157$  with Yates correction, n.s.), 'other' ( $\chi^2 = 1.4977$  with Yates correction, n.s.). The same goes for the case of collapsing overt pronouns and 'other' ( $\chi^2 = 0.0844$  with Yates correction, n.s.). **Figure 1** reports the comparisons.<sup>17</sup>

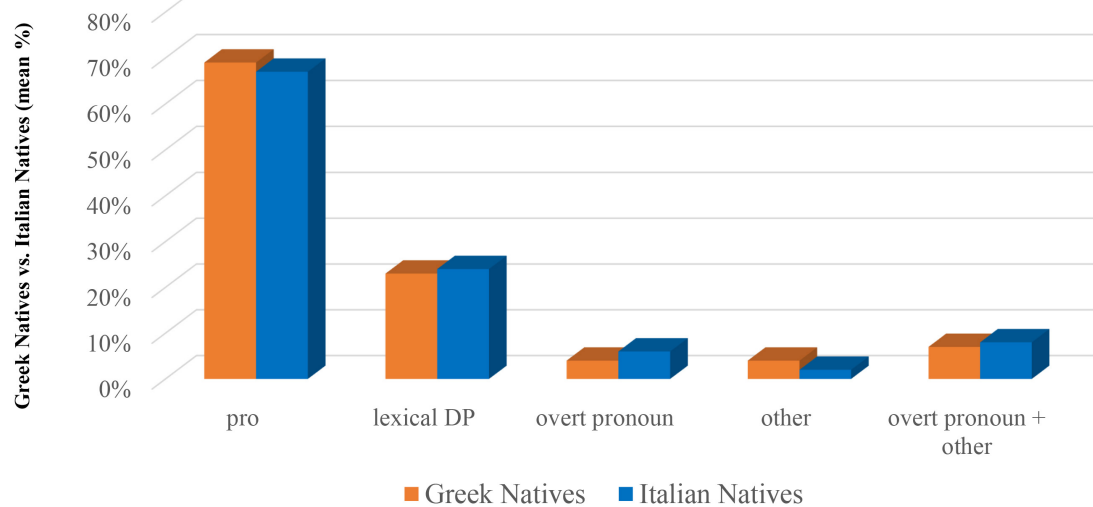
## Discussion

Results show a very similar pattern characterizing Italian native speakers' and Greek native speakers' choice of referring expressions. In particular, they show that there are no significant differences in the amounts of the various referring expressions chosen by the speakers. Null pronouns are widely employed, followed by lexical DPs, while overt pronouns are quite rare in both groups. Results are important in that they show that Italian and Greek, despite their differences, are comparable languages, at least as far as production is concerned, with respect to the relative amount of anaphoric devices employed.<sup>18</sup> This in turn means that in bilingual speakers of both these languages, no effect related to cross-linguistic influence is expected with respect to the issue at stake. With this in mind, we move to Study 2.

<sup>16</sup>Results are summarized in **Supplementary Table 1** (Italian Natives) and **Supplementary Table 2** (Greek Natives), where the Reference Total of the sentences for the two groups is shown, together with the indication of the clausal type and of the occurrences and percentages of the kind of referring expression employed.

<sup>17</sup>Given the small-scale nature of the data discussed in this study, we chose the  $\chi^2$ -test as a suitable non-parametric procedure to analyze our data. Group responses are indeed quite representative of individual ones, as revealed by a  $\leq 0.5$  coefficient of variation in responses for *pro*, lexical DPs and overt pronouns in Greek natives and for *pro* and lexical DPs in Italian natives. The latter holds for all the experimental groups discussed in the present work.

<sup>18</sup>As noted in footnote 1, Greek allows post-verbal subjects more than Italian. Our data support this fact in that post-verbal subjects in the Reference Total of the Greek natives are much more widespread (50.35%) than in the Reference Total of the Italian natives (21.42%). The difference is highly significant ( $\chi^2 = 22.6082$  with Yates correction, significant at  $p < 0.05$ , 0.01, and 0.005).



**FIGURE 1 |** Subject anaphoric devices in Greek Natives and Italian Natives.

## STUDY 2: SUBJECT ANAPHORIC DEVICES IN NATIVE AND NEAR-NATIVE SPEAKERS OF ITALIAN

The results of Study 1 show that native speakers of Italian and native speakers of Greek do not differ significantly in the production of null and overt pronominal as well as lexical DP subjects. Thus, we do not expect any effects of cross-linguistic influence with respect to the anaphoric devices chosen by the speakers of both these languages. These data will be relevant to establish whether the over-use of overt pronouns observed in near-natives by the studies described in the Introduction is due to cross-linguistic influence alone, or whether other factors are involved as well: if Greek-Italian bilingual speakers over-use overt pronouns, this cannot be due to cross-linguistic influence.

### Subjects

30 subjects participated in Study 2: the group of 10 native speakers of Italian of Study 1 (henceforth Natives), a group of 10 Greek-Italian bilinguals from birth living in Greece (henceforth Bilinguals in Greece), and a group of 10 native speakers of Greek who started to learn Italian after puberty and had reached a near-native level of proficiency in this language (henceforth L2ers).

Natives have been described in the section ‘Subjects’ of Study 1. As for Bilinguals in Greece (3 male; 7 female) their mean age at the time of testing was 21 (range 16–33). They were living in Greece at the time of testing and had been living there most of their lives. They were tested in Greece. They were all bilinguals from birth, with one parent native speaker of Greek and one parent native speaker of Italian. Despite living in Greece, they all also used Italian on a regular basis.<sup>19</sup> As for their education, 6 of

them were attending the last year of the Italian State School of Athens, 1 had just graduated from this school, 3 had a university degree, and had previously attended the Italian State School of Athens.

As for L2ers (4 male; 6 female), their mean age at the time of testing was 32 (range 21–52). They were born in Greece and had spent there at least the first 18 years of their lives. At the time of testing they were living in Italy, where they were tested. The length of their residence in Italy was 7 years on average. Their age of onset of exposure to Italian ranged from 15 to 28. As for their education 4 had a university degree and 6 had a high school degree and were attending university in Italy.

## Materials and Methods

### Ethical Considerations

The same ethical considerations holding for Study 1 (see section ‘Ethical Considerations’) hold for this study as well. The data collection at the Italian State School of Athens (which concerns 6 subjects, see section ‘Subjects’ above) was authorized by the school pro-Rector.

### Procedure

As described for Study 1, subjects were asked to watch The Pear Film and then tell the story, first in Italian and then in Greek. The subjects productions were recorded and then transcribed with the help of the CLAN system. Subjects were tested individually in a quiet room and the interviewer did not interact with them during their narration.

### The Near-Nativeness Level of the Subjects

In order to see whether the materials collected were appropriate for our study, we first performed a near-nativeness test on these materials, adapting White and Genesee’s (1996) near-nativeness test along the lines of Contemori et al. (2015) and Dal Pozzo and Matteini (2015). Three native speakers of Italian evaluated

<sup>19</sup>See section ‘The Near-Nativeness Level of the Subjects’ for more information concerning their level of proficiency in the two languages.

the oral productions in Italian of the experimental subjects, indicating their judgments with respect to five distinct aspects (morphology, syntax, vocabulary, pronunciation, fluency) on a scale of 10 cm.<sup>20</sup> The mean value of these five judgements constitutes the near-nativeness value assigned by each judge to each participant. The final near-nativeness value of each participant corresponds to the mean value of the values expressed by each judge. A speaker is considered near-native if her/his mean value ranges from 8.5 to 9.5.

Taken as a group, Bilinguals in Greece had a mean value of 8.98 (range 8.70–9.28). L2ers had a mean value of 8.88 (range 8.50–9.33). In order to have a line of comparison for our study, we had the same three judges evaluate the Natives productions as well: taken as a group, Natives had a mean value of 9.79 (range 9.64–9.96).

Although not entirely relevant for this study (but see section “Extension” below), we also asked three native speakers of Greek to evaluate the productions of the Bilinguals and the L2ers in Greek.<sup>21</sup> Taken as a group, Bilinguals had a mean value of 9.34 (range 8.61–9.80) while L2ers had a mean value of 9.73 (9.56–9.92). Note that the same Greek judges evaluated the productions of the group of the Greek native speakers of Study 1. Taken as a group, they had a mean value of 9.87 (range 9.75–10).<sup>22</sup>

### Defining the Reference Total

The Reference Total was derived with the same procedure described for Study 1. As mentioned, the Natives’ Reference Total consists of 387 sentences. The Bilinguals in Greece Reference Total consists of 241 sentences, while the L2ers’ Reference Total consists of 255 sentences.

## Results<sup>23</sup>

As in Study 1, *pro* is the preferred anaphoric device in all groups (67.18% Natives, 63.90% Bilinguals, 60.68% L2ers), followed by lexical DPs (24.28% Natives, 29.46% Bilinguals, 23.52% L2ers), overt pronouns (6.20% Natives, 5.80% Bilinguals, 14.50% L2ers) and ‘other’ (2.06% Natives, 0.82% Bilinguals, 1.17% L2ers).

As for *pro*, Natives do not differ from Bilinguals ( $\chi^2 = 0.7126$ , n.s.) nor from L2ers ( $\chi^2 = 2.7540$ , n.s.); Bilinguals and L2ers do not differ from each-other ( $\chi^2 = 0.5122$ , n.s.).

Lexical DPs as well appear equally employed: Natives do not differ from Bilinguals ( $\chi^2 = 2.0502$ , n.s.) nor from L2ers ( $\chi^2 = 0.0487$ , n.s.), Bilinguals and L2ers do not differ from each-other ( $\chi^2 = 2.2426$ , n.s.).

Similarly, as to the category ‘other,’ Natives do not differ from Bilinguals ( $\chi^2 = 0.7688$  with Yates correction, n.s.) nor from

L2ers ( $\chi^2 = 0.2918$  with Yates correction, n.s.), Bilinguals and L2ers do not differ from each-other ( $\chi^2 = 0.0040$  with Yates correction, n.s.).

Things appear different as far as overt pronouns are concerned. Natives do not differ from Bilinguals ( $\chi^2 = 0.0008$  with Yates correction, n.s.) but they significantly differ from L2ers ( $\chi^2 = 11.3923$  with Yates correction, significant at  $p < 0.05$ ; 0.01; 0.005). L2ers also significantly differ from Bilinguals ( $\chi^2 = 9.2462$  with Yates correction, significant at  $p < 0.05$ ; 0.01; 0.005).

These differences are replicated when overt pronouns and ‘other’ are collapsed: Natives do not differ from Bilinguals ( $\chi^2 = 0.3518$  with Yates correction, n.s.) but they significantly differ from L2ers ( $\chi^2 = 7.7651$  with Yates correction, significant at  $p < 0.05$ ; 0.01; L2ers significantly differ from Bilinguals ( $\chi^2 = 9.2427$  with Yates correction, significant at  $p < 0.05$ ; 0.01; 0.005). Results are shown in **Figure 2**.

## Discussion

Results clearly reveal that L2ers use significantly more overt pronouns than Natives and Bilinguals, while Bilinguals from birth behave like Natives in this respect. A significant difference between L2ers on one side and Natives and Bilinguals on the other is observed only with respect to overt pronouns (considered individually or collapsed with ‘other’).

Given that no effect related to cross-linguistic influence can be called into question in this respect for our experimental subjects (as revealed by Study 1), and that Bilinguals and L2ers have a comparable level of proficiency in Italian as attested, the relevant factor that Study 2 singles out is age of onset of exposure to Italian.<sup>24</sup>

Study 2 thus reveals first of all that over-use of overt subject pronouns also occurs in the absence of cross-linguistic influence. Furthermore, Study 2 reveals that it occurs only in a specific group of near-natives: i.e., only in those who have started to acquire the language in question after puberty. A further confirmation of this result is given in the following section.<sup>25</sup>

## Extension

In order to be sure that the results were not a by-product of a ‘stylistic choice’ made by these specific speakers, we compared L2ers productions in Italian with their productions in Greek. If the difference is maintained, it cannot be due to a personal stylistic choice of those speakers, otherwise we should find it also in their Greek productions. As we have shown in the section “The Near-Nativeness Level of the Subjects,” Greek is these subjects’ L1, and, despite their residence in Italy, they have preserved a native level of proficiency in this language (mean value 9.73,

<sup>20</sup>Two of the Italian judges (2 male; 1 female, aged 25–29, living in Italy) were teachers of Italian as an L2, and another was working for an organization for immigrants.

<sup>21</sup>The Greek judges (1 male; 2 female, aged 25–31, living in Greece) were teachers of Greek as an L2.

<sup>22</sup>Supplementary Table 9 reports the mean value of (near-) nativeness for each group of experimental subjects participating in Study 1, Study 2 and Study 3.

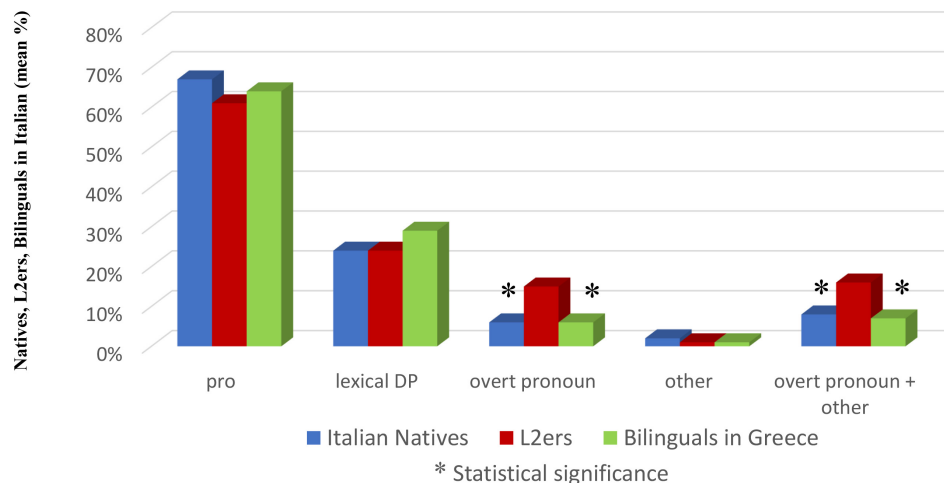
<sup>23</sup>Supplementary Table 3 reports the Reference Total concerning Bilinguals in Greece, together with the indication of the clausal type and of the occurrences and percentages of the kind of referring expression employed. The same is shown in Supplementary Table 4 for the L2ers. As for Natives, as already presented, the same is shown in Supplementary Table 1.

<sup>24</sup>Given the differences between Greek and Italian outlined in footnotes 1 and 18, we could expect cross-linguistic influence from Greek to Italian with respect to post-verbal subjects and the use of demonstratives for our L2ers.

This is however not the case: subjects in post-verbal position are 15% in the L2ers Reference Total (even less than in the Italian natives Reference Total) while demonstratives amount to only 10.8% of overt pronominal devices (8.33% in the Italian natives Reference Total).

<sup>25</sup>These results are also strengthened, as we shall see, by the findings in Study 3, where the productions of another group of bilinguals from birth (Bilinguals living in Italy) are analyzed. These bilinguals too, do not over-use overt pronouns in either of their languages.





**FIGURE 2 |** Subject anaphoric devices in Italian: Natives, L2ers and Bilinguals.

range 9.56–9.92). With the same procedure described for Study 1, we collected the materials and derived the Reference Total, consisting of 362 sentences (**Supplementary Table 5**).

Again, *pro* is the preferred anaphoric device (65.46%) followed by lexical DPs (27.34%), overt pronouns (4.69%) and ‘other’ (2.48%).

When we compared the L2ers productions in Greek with their productions in Italian, we found that there are no significant differences with regard to the implementation of *pro* ( $\chi^2 = 1.4176$ , n.s.), lexical DPs ( $\chi^2 = 1.1405$ , n.s.) and ‘other’ ( $\chi^2 = 0.7466$  with Yates correction, n.s.). There is, however, a significant difference for overt pronouns: these are attested to a significantly higher extent in Italian than in Greek ( $\chi^2 = 16.8345$  with Yates correction, significant at  $p < 0.05$ ; 0.01; 0.005). This significant difference is maintained when overt pronouns and ‘other’ are collapsed ( $\chi^2 = 10.4534$  with Yates correction, significant at  $p < 0.05$ ; 0.01; 0.005). This is shown in **Figure 3**.

As a final point, we compared the Greek productions of the L2ers with the Greek productions of the Greek native speakers of Study 1. No significant differences are attested: *pro* ( $\chi^2 = 1.4095$ , n.s.), lexical DP ( $\chi^2 = 1.9132$ , n.s.), ‘other’ ( $\chi^2 = 0.6661$  with Yates correction, n.s.), overt pronoun ( $\chi^2 = 0.2495$  with Yates correction, n.s.), overt pronouns and ‘other’ ( $\chi^2 = 0.0010$  with Yates correction, n.s.). L2ers over-use overt pronouns in their L2 only, while in their L1 their productions do not differ from those of other native speakers. This is shown in **Figure 4**.

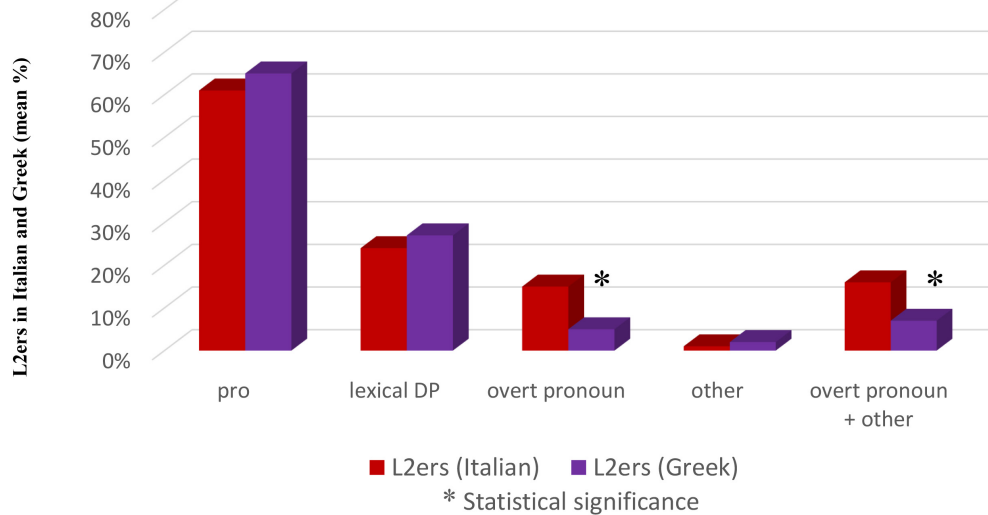
## Interim Conclusion

In the Introduction, we have briefly reviewed a number of studies that highlighted the role of cross-linguistic influence in determining over-use and over-acceptance of overt subject pronouns in co-reference with a topical antecedent in adult attrited speakers (Tsimpili et al., 2004), adult late acquirers (Sorace and Filiaci, 2006; Belletti et al., 2007) simultaneous bilingual children of a null and a non-null subject language (Serratrice et al., 2004; Sorace et al., 2009, a.o.). As these studies reveal, cross-linguistic influence seems to spread over

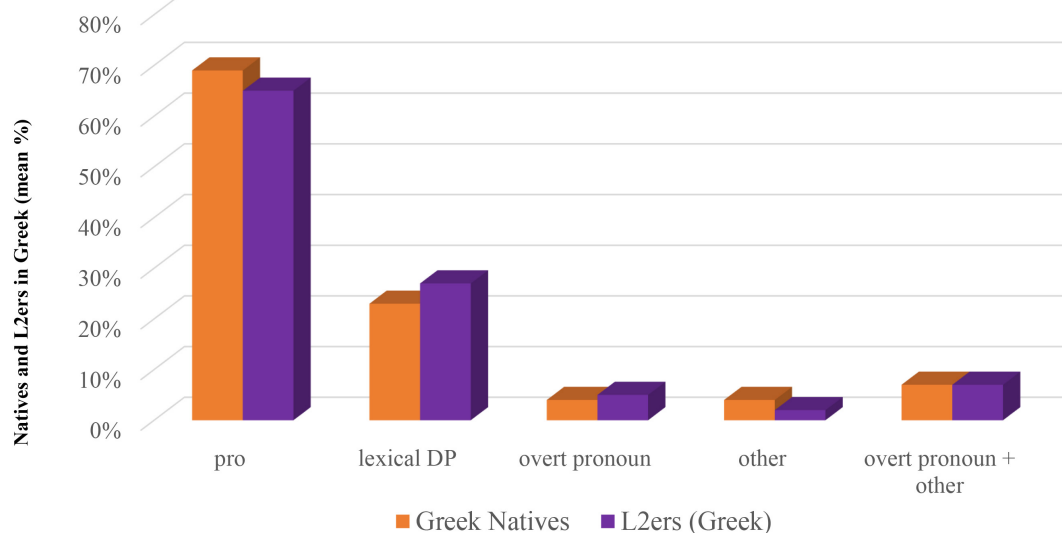
different populations of bilingual speakers, although in bilingual children developmental factors can be assumed to co-occur in determining its effects, as the results in Sorace et al. (2009) show. Particularly revealing in this respect are the differences between younger and older bilingual children (with the former choosing more overt pronouns), and those concerning monolingual children and monolingual adults (again, with the former choosing more overt pronouns). This fact, together with the observed directionality of cross-linguistic influence (from the non-null subject language to the null-subject language, but not the reverse) suggests that overt pronouns are somehow simpler than null ones. The results of Sorace et al. (2009) together with those of Filiaci et al. (2013) suggest on one side that not all null-subject languages are alike with respect to the division of labor between null and overt subject pronouns, and that cross-linguistic influence may occur also in bilinguals of two null-subject languages (as highlighted by Bini’s (1993) data as well).

Greek and Italian, as Study 1 reveals, are two null subject languages for which no significant quantitative differences are observed in the use of null subject pronouns, overt subject pronouns and subject lexical DPs, so that the results of Study 2 are not an effect of cross-linguistic influence. Here, we can see what cross-linguistic influence seems to obscure, i.e., a difference among different populations of near-natives, which singles out L2ers from bilinguals from birth. Another fact that Study 2 reveals is that, whatever the reason, on which we will not speculate in this work, overt pronouns appear simpler not only for children (as revealed by some of the studies quoted above) but also for adults, when age of onset of exposure to the language in question is rather late.<sup>26</sup>

<sup>26</sup>A brief examination of the contexts in which overt pronouns occurred reveals that while native speakers and bilinguals use them in topic shift contexts, in L2ers’ productions this is often not the case, especially when more than one Discourse Referent is active at some specific points of the narration. Overt pronouns were however very few in our corpora, and the issue needs to be studied more in depth and with a wider range of data. We therefore leave the issue for future research.



**FIGURE 3 |** Subject anaphoric devices in L2ers: Italian and Greek.



**FIGURE 4 |** Subject anaphoric devices in Greek: Natives and L2ers.

Absence of cross-linguistic influence has proved thus to offer a fruitful opportunity to study the role of other factors (e.g., age of onset of exposure): with this in mind, we move to Study 3.

### STUDY 3: THE ROLE OF DOMINANCE: COMPARING TWO GROUPS OF BILINGUALS

The results of Study 2 suggest that age of onset of exposure to Italian is a relevant factor in determining the over-use of overt pronouns in near-natives of Italian in the absence of effects related to cross-linguistic influence. Note that the two groups

were comparable, despite smaller, non-significant differences, as to the level of proficiency: they were both near-natives, and our aim is to compare natives and near-natives.<sup>27</sup>

<sup>27</sup> L2ers had a mean value of 8.88/10 in Italian and of 9.73/10 in Greek. Bilinguals in Greece had a mean value of 8.98/10 in Italian and of 9.34/10 in Greek. These differences (either within-group or between-group) are non-significant: L2ers Italian vs. Greek:  $\chi^2 = 0.0174$  with Yates correction, n.s.; Bilinguals in Greece: Greek vs. Italian  $\chi^2 = 0.2662$  with Yates correction, n.s.; L2ers vs. Bilinguals in Greece, Italian:  $\chi^2 = 0.4239$ , with Yates correction, n.s.; L2ers vs. Bilinguals in Greece, Greek:  $\chi^2 = 0.4196$  with Yates correction, n.s. The data concerning Bilinguals in Greece raise an interesting issue that we leave for future research, since Bilinguals in Greece appear to be near-natives in both Italian and Greek, i.e., they appear as native speakers of none of their two languages.

Level of proficiency, however, is not the only factor characterizing dominance, and if we want to study the role of dominance in near-natives, other factors have to be taken into consideration.<sup>28</sup>

In order to verify the role of dominance, we decided to compare two different groups of bilinguals: the bilinguals of Study 2, who were living in Greece (Bilinguals in Greece) and a group of bilinguals living in Italy (Bilinguals in Italy). Besides small, non-significant, differences concerning proficiency, the two groups differ in the dimension that concerns the language of the environment (or 'predominant' language, see Silva-Corvalán and Treffers-Daller, 2016:3).<sup>29</sup> Another relevant difference between the two groups concerns use: while Bilinguals in Greece use both Greek and Italian in everyday life, Bilinguals in Italy only use Italian in everyday life, reserving Greek basically for contacts with their family in Greece.

We first compared the Greek of these two groups, and then their Italian. Finally, an interesting comparison is a within-group comparison: the Greek vs. Italian of Bilinguals in Italy as well as the Greek vs. Italian of Bilinguals in Greece.

## Subjects

20 subjects participated in Study 3: the group of 10 Bilinguals living in Greece who participated in Study 2 (Bilinguals in Greece), and a group of 10 bilinguals living in Italy (henceforth Bilinguals in Italy). Bilinguals in Greece have already been described in Study 2. Bilinguals in Italy (4 male; 6 female) had a mean age of 22 (range 19–30). They had all been exposed to both languages since birth, with one parent native speaker of Greek and one parent native speaker of Italian. They grew up mostly in Greece (where they had all attended the Italian State School of Athens) and then they moved to Italy. Their residence in Italy was 6 years on average at the time of testing. As for education, 7 had a high school degree (and were attending university in Italy) and 3 had a university degree (taken in Italy). They were tested in Italy.

## Materials and Methods

### Ethical Considerations

The same ethical considerations holding for Study 1 (see section "Ethical Considerations") and Study 2 (see section "Ethical Considerations") hold here as well.

<sup>28</sup>See a.o. Birdsong (2014), Montrul (2016), Silva-Corvalán and Treffers-Daller (2016), Treffers-Daller and Korybski (2016) and the references quoted there.

<sup>29</sup>Bilinguals in Italy had a mean value in Italian of 9.03 (range 8.69–9.38) and of 8.79 in Greek (range 8.08–9.24). As we noticed for the Bilinguals in Greece, Bilinguals in Italy appear to be native speakers of neither of their languages, too. Another fact worth noting is that their residence in Italy seems to have had a greater effect on their Greek, when compared to L2ers, which still maintain their native level in Greek notwithstanding the years spent in Italy: attrition seems to have a more pervasive effect in bilinguals than in L2ers. The near-nativeness values in each language of this group of experimental subjects is also reported in **Supplementary Table 9**. As for within-group and between-group statistical significance, we observed the following: Bilinguals in Italy: Greek vs. Italian  $\chi^2 = 0.2974$  with Yates correction, n.s.; Bilinguals in Greece vs. Bilinguals in Italy: Greek  $\chi^2 = 0.1195$  with Yates correction, n.s.; Italian  $\chi^2 = 0.5036$  with Yates correction, n.s.

## Procedure

The procedure employed to collect the data is the same described for Study 1 and Study 2, the only difference being that Bilinguals in Italy first used Greek and then Italian to tell the story. The procedure to analyze data (sentence typing, derivation of the Reference Total, determination of the subjects' near-nativeness value) is the same described for Study 1 and Study 2, as well.

## Results<sup>30</sup>

The Reference Total concerning the Greek of Bilinguals in Italy consists of 251 sentences, while for their Italian of 234 sentences. The Reference Total of the Greek of Bilinguals in Greece consists of 267 sentences, while that of their Italian of 241 sentences as described in Study 2.

As mentioned, we will first perform a between-group comparison, initially comparing the Greek of the two groups, then their Italian. We will then proceed to a within-group comparison, first on Bilinguals in Italy, then on Bilinguals in Greece.

### Bilinguals in Italy vs. Bilinguals in Greece

#### *Bilinguals in Italy vs. Bilinguals in Greece: Greek*

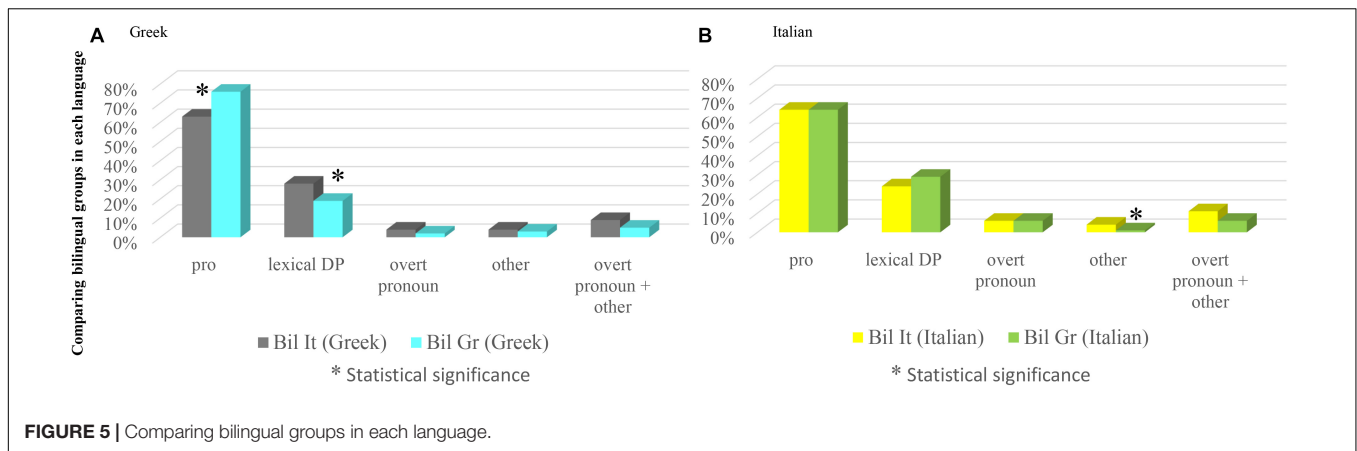
In both groups *pro* is the anaphoric device employed most (63.34% in Bilinguals in Italy; 76.02% in Bilinguals in Greece), followed by lexical DPs (27.88% Bilinguals in Italy; 19.10% Bilinguals in Greece), overt pronouns (4.38% Bilinguals in Italy; 2.24% Bilinguals in Greece) and 'other' (4.38% Bilinguals in Italy; 2.62% Bilinguals in Greece). When we compare the percentage rates, we do not find any significant difference concerning overt pronouns ( $\chi^2 = 1.2466$  with Yates correction, n.s.) or 'other' ( $\chi^2 = 0.7285$  with Yates correction, n.s.).<sup>31</sup> The employment of lexical DPs instead differs significantly ( $\chi^2 = 5.5802$ , significant at  $p < 0.05$ ), as well as use of *pro*, where the difference is highly significant ( $\chi^2 = 9.8889$ , significant at  $p < 0.05$ ; 0.01; 0.005). Bilinguals in Italy use significantly less *pro* and significantly more lexical DPs when compared to Bilinguals in Greece. Results are shown in **Figure 5A**.

#### *Bilinguals in Italy vs. Bilinguals in Greece: Italian*

In Italian *pro* is the mostly employed anaphoric device in both groups, too (64.52% Bilinguals in Italy; 63.90% Bilinguals in Greece), followed by lexical DPs (24.35% Bilinguals in Italy; 29.46% Bilinguals in Greece), overt pronouns (6.83% Bilinguals in Italy; 5.80% Bilinguals in Greece) and 'other' (4.27% Bilinguals in Italy; 0.82% Bilinguals in Greece). When we turn to the comparisons, we do not find any significant difference as far as overt pronouns are concerned ( $\chi^2 = 0.0740$  with Yates correction, n.s.), but we find a significant difference with respect to 'other'

<sup>30</sup>The Reference Total concerning the Greek of Bilinguals in Italy is reported in **Supplementary Table 6**, while the Reference Total concerning the Italian of Bilinguals in Italy is reported in **Supplementary Table 7**. **Supplementary Table 8** reports the Reference Total concerning the Greek of Bilinguals in Greece, while the Reference Total concerning their Italian is shown in **Supplementary Table 3**. The tables show the Reference Total of the sentences produced, together with the clausal type and the occurrences and percentages of the kind of referring expression employed.

<sup>31</sup>When pronouns and other are collapsed we do not reach significance either, as expected ( $\chi^2 = 2.5295$  with Yates correction, n.s.).



( $\chi^2 = 4.4045$  with Yates correction, significant at  $p < 0.05$ ). The difference doesn't reach significance when overt pronouns and 'other' are collapsed ( $\chi^2 = 2.4172$  with Yates correction, n.s.). We do not find any significant difference with respect to *pro* ( $\chi^2 = 0.0205$ , n.s.) or lexical DPs ( $\chi^2 = 1.5696$ , n.s.). Results are shown in Figure 5B.

### Interim discussion

As Figure 5 shows, the Italian of these two groups of speakers is quite uniform, with the exception of a significant difference concerning 'other,' more employed by Bilinguals in Italy.

There are indeed some interesting differences concerning the Greek of these two groups of speakers, in that, compared to Bilinguals in Greece, Bilinguals in Italy use significantly less *pro* and significantly more lexical DPs. This could *prima facie* suggest that, although dominance does not affect the productions of overt pronouns, it has some effects in the choice of referring expressions, in that *pro* is less used by those speakers who don't use this language in everyday life. This conclusion, however, needs further confirmation, since it could be rather the group which uses both languages in everyday life, i.e., Bilinguals in Greece, the one who manifests a peculiarity.

### Within-Group Comparison

#### Bilinguals in Italy: Greek vs. Italian

The within-group comparison concerning the Greek and the Italian of Bilinguals in Italy shows that *pro* is the most employed anaphoric device in both languages (63.34% in Greek; 64.52% in Italian) followed by lexical DPs (27.88% in Greek; 24.35% in Italian), overt pronouns (4.38% in Greek; 6.83% in Italian) and other (4.38% in Greek; 4.27% in Italian). The comparison reveals no significant differences with respect to *pro* ( $\chi^2 = 0.0735$ , n.s.), lexical DP ( $\chi^2 = 0.7805$ , n.s.), overt pronouns ( $\chi^2 = 0.9608$ , n.s.), 'other' ( $\chi^2 = 0.0270$ , n.s.), nor when collapsing overt pronouns and 'other' ( $\chi^2 = 0.5076$ , n.s.). Results are shown in Figure 6A.

#### Bilinguals in Greece: Greek vs. Italian

The within-group comparison concerning the Greek and the Italian of Bilinguals in Greece shows that *pro* is the most

employed anaphoric device in both languages (76.02% in Greek; 63.90% in Italian), followed by lexical DPs (19.10% in Greek; 29.46% in Italian), overt pronouns (2.24% in Greek; 5.80% in Italian), and 'other' (2.62% in Greek; 0.82% in Italian). The comparison reveals, however, some significant differences: this is so in the case of *pro* ( $\chi^2 = 8.9215$ , significant at  $p < 0.05$ ; 0.01; 0.005) and of lexical DPs ( $\chi^2 = 7.4494$ , significant at  $p < 0.05$ ; 0.01). The use of overt pronouns instead does not reveal significant differences in the two languages ( $\chi^2 = 3.3596$  with Yates correction, n.s.), as well as 'other' ( $\chi^2 = 1.4207$  with Yates correction, n.s.) or, as expected, collapsing overt pronouns and 'other' ( $\chi^2 = 0.4451$  with Yates correction, n.s.). Results are shown in Figure 6B.

### Interim discussion

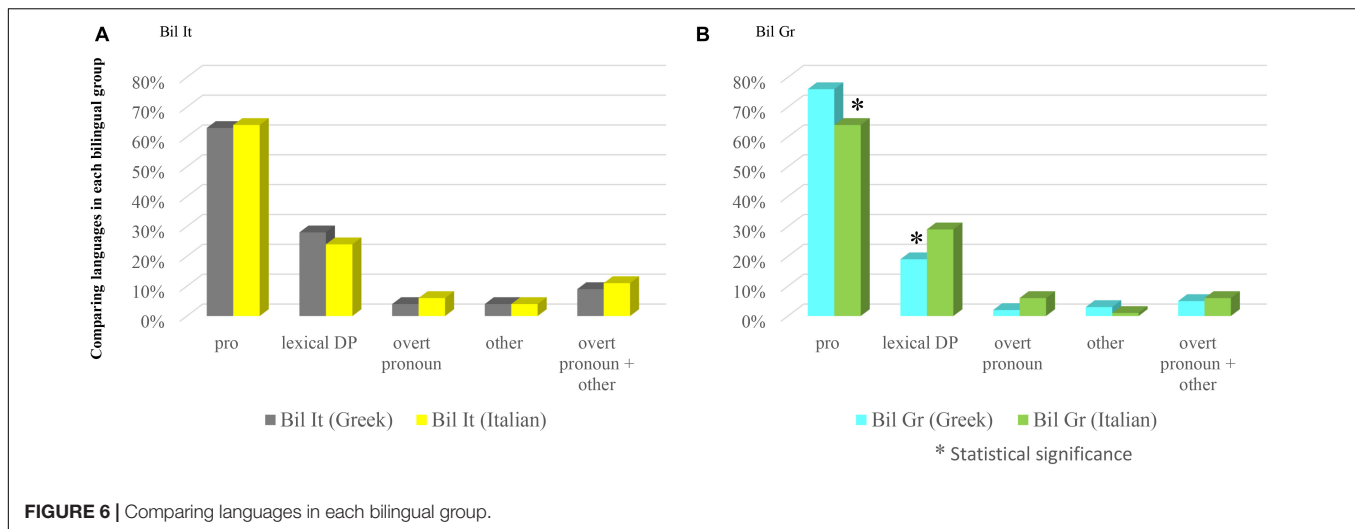
The within-group comparison, shown in Figure 6, reveals an interesting fact: while Bilinguals in Italy make the same choice of referential expressions in the language they daily use (Italian) and in the one they seldom use (Greek), Bilinguals in Greece instead differ significantly: they use significantly more *pro* in Greek than in Italian, conversely using more lexical DPs in Italian than in Greek. In contrast, overt pronouns are used to a comparable extent in the two languages.

As we have seen in Study 1, however, in Italian we did not find significant differences between Bilinguals in Greece and Native speakers of Italian, neither for overt pronouns, nor *pro*, nor lexical DPs. There are therefore strong reasons to believe that the difference concerns rather their predominant language, i.e., Greek.

### Extensions and Final Discussion

At this point, in order to have a clearer picture, we will compare the Greek of all groups discussed in this paper (Natives, Bilinguals in Greece, L2ers, Bilinguals in Italy) as well as their Italian. Let's start with Greek. As for *pro*, Bilinguals in Greece significantly differ from Bilinguals in Italy (as shown in the section 'Bilinguals in Italy vs. Bilinguals in Greece: Greek') and from L2ers ( $\chi^2 = 8.1529$ , significant at  $p < 0.05$ ; 0.01) though not from Natives ( $\chi^2 = 3.6719$ , n.s.). As for lexical DPs, Bilinguals in Greece again, besides the significant difference with respect to Bilinguals in Italy singled out in the section 'Bilinguals in Italy

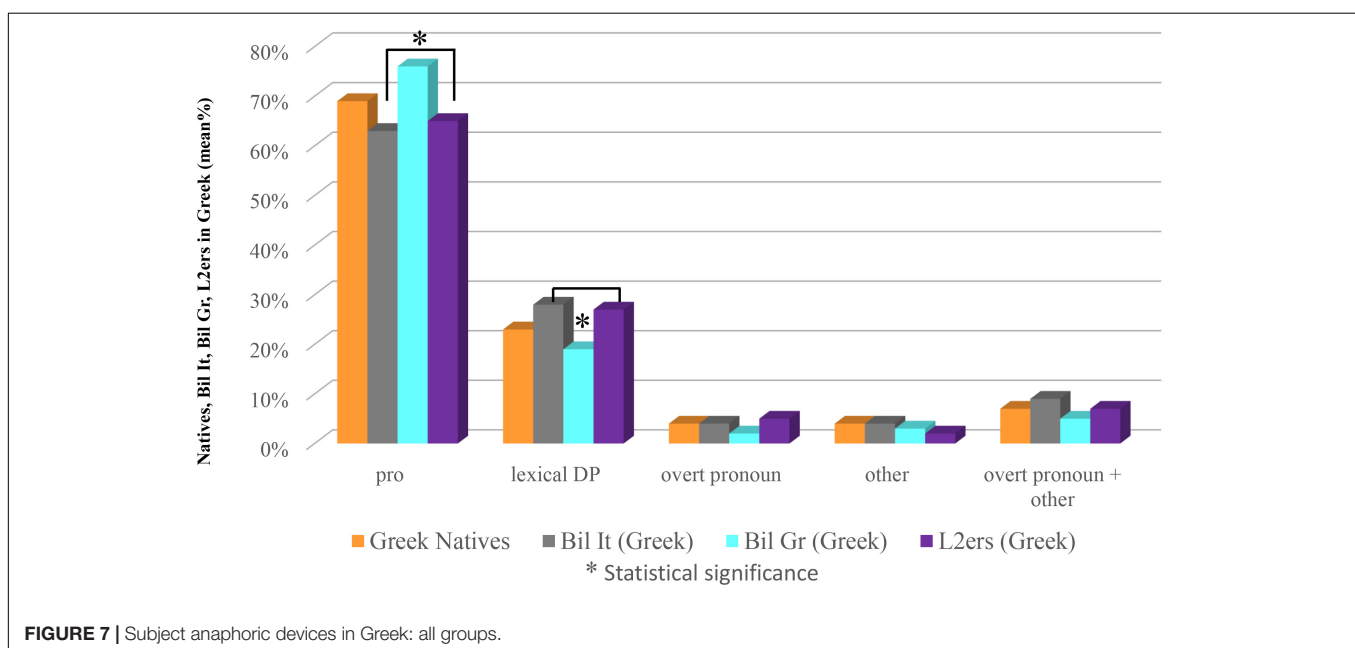




vs. Bilinguals in Greece: Greek') show a significant difference also with respect to L2ers ( $\chi^2 = 5.7548$ , significant at  $p < 0.5$ ), though not with respect to Natives ( $\chi^2 = 1.6077$ , n.s.). We didn't find any other significant difference in this comparison: for *pro*, Bilinguals in Italy vs. Natives ( $\chi^2 = 2.6737$ , n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 0.2921$ , n.s.); for lexical DPs, Bilinguals in Italy vs. Natives ( $\chi^2 = 1.9631$ , n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 0.0217$ , n.s.); for overt pronouns, Bilinguals in Italy vs. Natives ( $\chi^2 = 0.0458$  with Yates correction, n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 0.0002$  with Yates correction, n.s.), Bilinguals in Greece vs. Natives ( $\chi^2 = 0.7838$  with Yates correction, n.s.), Bilinguals in Greece vs. L2ers ( $\chi^2 = 1.9670$  with Yates correction, n.s.); for 'other,' Bilinguals in Italy vs. Natives ( $\chi^2 = 0.0458$  with Yates correction, n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 1.1414$  with Yates correction, n.s.), Bilinguals in Greece vs. Natives ( $\chi^2 = 0.3559$  with Yates correction, n.s.), Bilinguals in Greece vs.

L2ers ( $\chi^2 = 0.0223$  with Yates correction, n.s.); for overt pronoun + 'other,' Bilinguals in Italy vs. Natives ( $\chi^2 = 0.2065$  with Yates correction, n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 0.3185$  with Yates correction, n.s.), Bilinguals in Greece vs. Natives ( $\chi^2 = 1.4884$  with Yates correction, n.s.), Bilinguals in Greece vs. L2ers ( $\chi^2 = 1.0442$  with Yates correction, n.s.). Results are shown in Figure 7.

As far as Italian is concerned results show that L2ers significantly differ in the use of overt pronouns not only with respect to Natives and Bilinguals in Greece (as shown in Study 2) but also with respect to Bilinguals in Italy ( $\chi^2 = 6.6599$  with Yates correction, significant at  $p < 0.05$ ; 0.01). This significance with respect to Bilinguals in Italy is lost when overt pronouns are collapsed with 'other' ( $\chi^2 = 1.8134$  with Yates correction, n.s.). As we have seen in the section "Bilinguals in Italy vs. Bilinguals in Greece: Italian," Bilinguals in Italy use significantly



more ‘other’ than Bilinguals in Greece. It is not so when we compare Bilinguals in Italy to L2ers ( $\chi^2 = 3.4052$  with Yates correction, n.s.) or to Natives ( $\chi^2 = 1.7991$  with Yates correction, n.s.). We didn’t find any other significant difference in this comparison: for *pro*, Bilinguals in Italy vs. Natives ( $\chi^2 = 0.4588$ , n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 0.7310$ , n.s.); for lexical DPs, Bilinguals in Italy vs. Natives ( $\chi^2 = 0.0004$ , n.s.), Bilinguals in Italy vs. L2ers ( $\chi^2 = 0.0461$ , n.s.); for overt pronoun, Bilinguals in Italy vs. Natives ( $\chi^2 = 0.0208$  with Yates correction, n.s.); for overt pronoun + ‘other,’ Bilinguals in Italy vs. Natives ( $\chi^2 = 1.0759$  with Yates correction, n.s.). Results are shown in **Figure 8**.

The comparisons shown in **Figure 7**, concerning Greek, single out that Bilinguals in Greece use significantly more *pros*, and significantly less lexical DPs, when compared to both Bilinguals in Italy and L2ers, though not when compared to Natives. The fact that the difference is not restricted to a single group, together with the data described in the section “Within-Group Comparison” allows us to argue that it is precisely this group of speakers that is doing something peculiar, and that is doing so in the predominant language. As we can see in **Figure 1** (which pertains to Study 1) native speakers of Greek use more *pros* and less lexical DPs than Italian natives. This difference, as we noted, is far from significant, however: it has led us to assume that Italian and Greek are very similar with respect to the choice of anaphoric devices. What these bilinguals do, we argue, is amplifying this little difference, modifying their choices in the predominant language. Similar facts have been noted in situations of language contact (see e.g., Scala, 2018 and the references quoted there), where two languages appear more divergent when they are in contact than when they are spoken in non-contact areas, and have been considered therefore a driving factor of language change. As we said in the section “Materials and Methods,” Bilinguals in Greece use Italian (as well as Greek) on a regular basis (differently from Bilinguals in Italy, as well as from L2ers, who use Greek

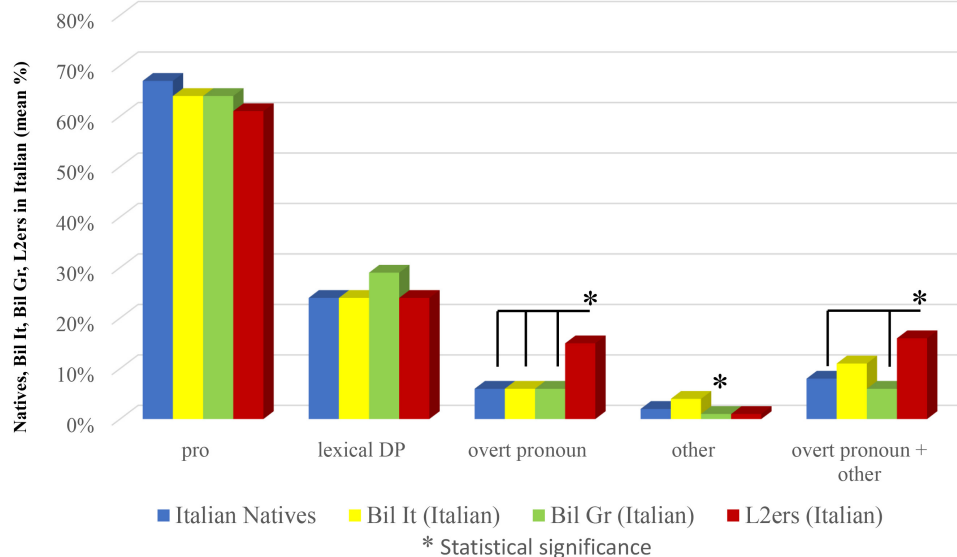
basically just for contacts with their family in Greece). Bilinguals in Greece either attend the Italian State School of Athens or use Italian for their work, and they live in Greece. They are the only group, among our experimental subjects, who employs the two languages in everyday life. Amplifying the differences in the two languages helps these bilingual speakers keeping the two languages separate.

Interestingly, the modification does not involve overt pronouns, but lexical DPs as well as null pronouns. This suggests that overt pronouns are a really marked option, questioning accessibility marking scales such as those in Ariel (1990, 2001) which place overt pronouns near to null ones.

The comparisons shown in **Figure 8**, concerning Italian, confirm the results of Study 2 (over-use of subject overt pronouns by L2ers) and extend their validity with respect to Bilinguals in Italy (though significance is lost when overt pronouns are collapsed with ‘other’). They also highlight that the significant over-use of ‘other’ by Bilinguals in Italy with respect to Bilinguals in Greece is restricted to this case, hence no reliable conclusions can be drawn in this respect.

## CONCLUSION

In Study 1, we have presented evidence that native speakers of Greek and of Italian do not differ significantly in the choice of subject anaphoric devices, at least as far as production is concerned: *pro* is overwhelmingly the most attested device, followed by lexical DPs, while overt pronouns are very few in both groups. Italian-Greek is therefore a suitable language combination if we want to study bilinguals’ choices in this respect, since effects related to cross-linguistic influence are absent. This does not mean, of course, that we want to deny, in general, the effects of cross-linguistic influence on the choice of anaphoric devices in bilinguals, since this is clearly demonstrated by several



**FIGURE 8 |** Subject anaphoric devices in Italian: all groups.

studies. Absence of cross-linguistic influence, however, allows the discovery of other factors playing a role in the issue at stake.

In Study 2, we have compared the productions in Italian of a group of native speakers and two groups of near-natives: a group of bilinguals from birth (Bilinguals in Greece) and a group with post-puberty age of onset of exposure to Italian (L2ers). We have given evidence that over-use of overt subject pronouns in near-natives of Italian takes place when effects related to cross-linguistic influence are absent, singling out that this holds for a specific population of near-natives: those with age of initial exposure to the language in question after puberty. Tsimpli (2014) argues that phenomena which are acquired late (such as pragmatically conditioned aspects of pronominal use) do not cause pronounced differences among bilinguals differing for age of initial exposure. Our study suggests that this claim is valid for pre-puberty but not for post-puberty age of onset of exposure.

As a reviewer wisely observes, the two groups of near-natives in Study 2 do not differ only with respect to age of onset of exposure to Italian, but also with respect to language of the environment: while L2ers live in Italy, Bilinguals in Greece live in Greece. The comparison between L2ers and another group of bilinguals from birth (the Bilinguals in Italy of Study 3) with the same language of the environment as the L2ers (Italian) confirms, however, the very same result: L2ers resort to overt pronouns significantly more than native speakers and bilinguals from birth, as shown in **Figure 8**.

The language of the environment (or ‘majority language,’ or ‘predominant language’), one of the variables characterizing dominance, does not seem to have an effect on the choice of overt pronouns, as confirmed by Study 3.

This variable, however, combined with regular use of the two languages, has indeed an effect in the choice of anaphoric devices such as *pro* and lexical DPs, though in a direction we did not expect: it is in the predominant language, rather than in the non-predominant one, that differences between natives and bilinguals have been observed. We have interpreted these differences as stemming from the bilinguals’ need to keep the two languages they daily use as distant as possible. Interestingly, these differences do not involve overt pronouns, but concern a wider use of null pronouns which charges lexical DPs. This suggests that overt pronouns are a marked option, questioning accessibility marking scales such as those in Ariel (1990, 2001) which place overt pronouns near to null ones. As a reviewer suggests, the significantly higher use of *pro* in Greek by Bilinguals in Greece might also reflect an underlying property of Greek *pro*, which, according to some authors appears to be compatible with salient/subject antecedent but also with non-salient/ object antecedent (Dimitriadis, 1996; Torregrossa et al., 2015). At a first analysis, our data are not very clear in this respect, and we have to leave this issue for future research.

## REFERENCES

Alexiadou, A., and Anagnostopoulou, E. (2002). “Raising without infinitives and the role of agreement,” in *Dimensions of Movement*, eds A. Alexiadou, E. Anagnostopoulou, S. Barriers, and H. M. Gärtner (Amsterdam: John Benjamins Publishing Company), 17–30.

A final note concerns the small-scale nature of our corpora, which has proven particularly limiting in the case of overt pronouns (which are seldom produced by our subjects) preventing a serious qualitative analysis of the contexts in which they occur in natives, bilinguals and L2ers. Another issue which we leave for future research is thus an inquiry with a wider range of data, collected with the help of different tasks, as a reviewer suggests.

## AUTHOR CONTRIBUTIONS

EDD developed the rationale of the study, the study concept and design, and wrote the manuscript. IB contributed in data collection. EDD and IB contributed in data analysis and its interpretation. Both authors critically read the manuscript providing comments that helped to improve its final version, and approved the final version for submission.

## FUNDING

This research was supported by the Università per Stranieri di Perugia, D.R. 196/2018 and Progetto di Ricerca di Ateneo Di Domenico 2017 and Di Domenico 2018.

## ACKNOWLEDGMENTS

Parts of this work have been presented at Eurosia 27 (University of Reading, August 29–September 2, 2017), at GALA 13 (Universitat de les Illes Balears September 7–9, 2017), at the Linguistics Colloquia held at the University of Perugia (November 21, 2017), at the 44th Incontro di Grammatica Generativa (Università di Roma III, March 1–3, 2018) and at the workshop ‘Overt subject pronouns in null-subject languages’ (Università per Stranieri di Perugia, September 13, 2018). We thank the audience at these conferences and the two Frontiers reviewers of this paper for helpful comments and suggestions. Special thanks are due to the Italian State School of Athens (in the person of Marta Zanardo, pro-Rector of the school at the time of the data collection), and to all our experimental subjects. Finally, we would like to thank Antonello Belli, Carla Contemori, Francesca Filiaci, Jihye Kang, and Simona Matteini. All errors and shortcomings are of course our own.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02729/full#supplementary-material>

Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. London: Routledge.

Ariel, M. (2001). “Accessibility theory: an overview,” in *Text Representation: Linguistic and Psycholinguistic aspects*, eds T. Sanders, J. Schilperoord, and W. Spooren (Amsterdam: John Benjamins Publishing Company), 29–87. doi: 10.1075/hcp.8.04ari

- Belletti, A. (2001). "Inversion' as focalization," in *Subject Inversion and the Theory of Universal Grammar*, eds A. Hulk and J. Y. Pollock (Oxford: Oxford University Press), 60–90.
- Belletti, A. (2004). "Aspects of the low IP area," in *The Cartography of Syntactic Structures vol. II: The Structure of CP and IP*, ed. L. Rizzi (Oxford: Oxford University Press), 16–51.
- 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
- Bini, M. (1993). "La adquisición del italiano: mas allá de las propiedades sintácticas del parámetro pro-drop," in *La Linguística y el Analisis de los Sistemas no Natives*, ed. L. Liceras (Ottawa: Doverhouse), 126–139.
- Birdsong, D. (2014). Dominance and age in bilingualism. *Appl. Linguist.* 35, 374–392. doi: 10.1093/applin/amu031
- Calabrese, A. (1986). "Pronomina," in *Papers in Theoretical Linguistics*, eds N. Fukui, T. R. Rapoport, and E. Sagey (Cambridge, MA: MITWPL), 1–46.
- Carminati, M. N. (2002). *The Processing of Italian Subject Pronouns*. Doctoral Dissertation, University of Massachusetts, Amherst, MA.
- Chafe, W. (ed.) (1980). *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.
- Chiou, M. (2013). Performing anaphora in modern greek: a neo-gricean pragmatic analysis. *Res. Lang.* 11, 335–358. doi: 10.2478/v10015-012-0029-1
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Contemori, C., Dal Pozzo, L., and Matteini, S. (2015). "Resolving pronominal anaphora in real-time: a comparison between Italian native and near-native speakers," in *Structures, Strategies and Beyond. Studies in Honour of Adriana Belletti*, eds E. Di Domenico, C. Hamann, and S. Matteini (Philadelphia, PA: John Benjamins Publishing Company), 257–276. doi: 10.1075/la.223.12con
- Dal Pozzo, L., and Matteini, S. (2015). "(In)definiteness and near-nativeness: article choice at interfaces," in *Language Acquisition and Development: Proceedings of GALA 2013*, eds C. Hamann and E. Ruigendijk (Newcastle upon Tyne: Cambridge Scholars Publishing), 61–82.
- Dimitriadis, A. (1996). "When pro-drop languages don't: overt pronominal subjects and pragmatic inference," in *Proceedings of the Chicago Linguistic Society*, Vol. 32, eds L. M. Dobrin, K. Singer, and L. McNair (Chicago, IL: Chicago Linguistic Society), 33–47.
- Filiaci, F., Sorace, A., and Carreiras, M. (2013). Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Lang. Cogn. Neurosci.* 28, 825–843. doi: 10.1080/01690965.2013.801502
- Grice, P. (1975). "Logic and conversation," in *Syntax and Semantics - Speech Acts*, eds P. Cole and J. L. Morgans (New York, NY: Academic Press), 41–58.
- Hacohen, A., and Schaeffer, J. (2007). Subject realization in early Hebrew-English bilingual acquisition: the role of cross-linguistic influence. *Bilingualism* 10, 333–344. doi: 10.1017/S1366728907003100
- Hulk, A., and Müller, N. (2000). Cross-linguistic influence at the interface between syntax and pragmatics. *Bilingualism* 3, 227–244. doi: 10.1017/S1366728900000353
- Ingria, R. (2005). Grammatical formatives in a generative lexical theory. the case of modern Greek καί. *J. Greek Linguist.* 6, 61–101. doi: 10.1075/jgl.6.06ing
- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620607
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Tool*, 3rd Edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Montrul, S. (2016). "Dominance and proficiency in early and late bilingualism," in *Language Dominance in Bilinguals. Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35. doi: 10.1017/CBO9781107375345.002
- Panagiotidis, P. (2000). Demonstrative determiners and operators: the case of Greek. *Lingua* 110, 717–742. doi: 10.1016/S0024-3841(00)00014-0
- Paradis, J., and Navarro, S. (2003). Subject realization and cross-linguistic interference in the bilingual acquisition of Spanish and English: what is the role of input? *J. Child Lang.* 30, 1–23. doi: 10.1017/S0305000903005609
- Philippaki-Warbuton, L., and Catsimali, G. (1999). "On control in modern Greek," in *Studies in Greek Syntax*, eds A. Alexiadou, G. Horrocks, and M. Stavrou (Dordrecht: Kluwer Academic Publishers), 153–168. doi: 10.1007/978-94-015-9177-5\_8
- Pinto, M. (2006). "Subject pronouns in bilinguals: interference or maturation?," in *The Acquisition of Syntax in Romance Languages*, eds V. Torrens and L. Escobar (Amsterdam: Benjamins), 331–352. doi: 10.1075/lald.41.16pin
- Rizzi, L. (1982). *Issues in Italian Syntax*. Berlin: Walter de Gruyter. doi: 10.1515/9783110883718
- Rizzi, L. (2005). "On some properties of subjects and topics," in *Proceedings of the XXX Incontro di Grammatica Generativa*, eds L. Brugè, G. Giusti, N. Munaro, W. Schweikert, and G. Turano (Venezia: Cafoscarina), 203–224.
- Rizzi, L. (2018). "Subjects, topics and the interpretation of pro," in *From Sounds to Structures: Beyond the Veil of Maya*, eds R. Petrosino, P. Cerrone, and H. van der Hulst (Berlin: De Gruyter Mouton), 510–529. doi: 10.1515/9781501506734-019
- Roussou, A., and Tsimpli, I. M. (2006). On Greek VSO again! *J. Linguist.* 42, 317–344. doi: 10.1017/S0022226706003914
- Scala, A. (2018). "I dialetti armeni tra continuità lessicale e innovazioni strutturali a modello turco," in *Mutamento Linguistico e Biodiversità. Atti del XVI Convegno Annuale Della Società Italiana di Glottologia*, eds L. Costamagna, E. Di Domenico, A. Marcaccio, S. Scaglione, and B. Turchetta (Roma: Il Calamo), 189–208.
- Serratrice, L., Sorace, A., and Paoli, S. (2004). Crosslinguistic influence at the syntax-pragmatics interface: subjects and objects in Italian-English bilingual and monolingual acquisition. *Bilingualism* 7, 183–206. doi: 10.1017/S1366728904001610
- Silva-Corvalán, C., and Treffers-Daller, J. (2016). "Digging into dominance: a closer look at language dominance in bilinguals," in *Language Dominance in Bilinguals. Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 1–14. doi: 10.1017/CBO9781107375345
- Sorace, A. (2006). "Interfaces in L2 development," in *Language Acquisition and Development. Proceedings of GALA 2005*, eds A. Belletti, E. Bennati, C. Chesì, E. Di Domenico, and I. Ferrari (Newcastle upon Tyne: Cambridge Scholars Press), 505–521.
- Sorace, A. (2011). Pinning down the concept of 'interface' in bilinguals. *Linguist. Approaches Biling.* 1, 1–33. doi: 10.1075/lab.1.1.01sor
- Sorace, A., and Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Lang. Res.* 22, 339–368. doi: 10.1191/0267658306sr271oa
- Sorace, A., Serratrice, L., Filiaci, F., and Baldo, M. (2009). Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children. *Lingua* 119, 460–477. doi: 10.1016/j.lingua.2008.09.008
- Spyropoulos, V. (2007). "Finiteness and control in Greek," in *New Horizons in the Analysis of Control and Raising*, eds W. D. Davies and S. Dubinsky (Dordrecht: Springer), 159–183. doi: 10.1007/978-1-4020-6176-9\_7
- Torregrossa, J., Bongartz, C., and Tsimpli, I. (2015). "Testing accessibility: a cross-linguistic comparison of the syntax of referring expressions," in *Proceedings of the 89th Annual Meeting of the Linguistic Society of America (Portland)*, (Salt Lake, UT: Linguistic Society of America). doi: 10.3765/exabs.v0i0.3046
- Treffers-Daller, J., and Korybski, T. (2016). "Using lexical diversity measures to operationalize language dominance in bilinguals," in *Language Dominance in Bilinguals. Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 106–133. doi: 10.1017/CBO9781107375345.006
- Tsimpli, I. (2014). Early, late or very late. Timing acquisition and bilingualism. *Linguist. Approaches Biling.* 4, 283–313. doi: 10.1075/lab.4.3.01tsi
- Tsimpli, I., Sorace, A., Heycock, C., and Filiaci, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near native speakers of English. *Int. J. Biling.* 8, 257–277. doi: 10.1177/13670069040080030601
- White, L., and Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Lang. Res.* 12, 233–265. doi: 10.1177/026765839601200301

**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 © 2019 Di Domenico and Baroncini. 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.





# Input Dominance and Development of Home Language in Russian-German Bilinguals

Natalia Gagarina<sup>1\*</sup> and Annegret Klassert<sup>2</sup>

<sup>1</sup> Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany, <sup>2</sup> University of Applied Science Clara Hoffbauer, Potsdam, Germany

## OPEN ACCESS

### Edited by:

Cornelia Hamann,  
University of Oldenburg, Germany

### Reviewed by:

Vicky Chondrogianni,  
University of Edinburgh,  
United Kingdom  
Tanja Anstatt,  
Ruhr-Universität Bochum, Germany

### \*Correspondence:

Natalia Gagarina  
gagarina@leibniz-zas.de

### Specialty section:

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

**Received:** 02 May 2018

**Accepted:** 28 August 2018

**Published:** 01 October 2018

### Citation:

Gagarina N and Klassert A (2018)  
Input Dominance and Development of  
Home Language in Russian-German  
Bilinguals. *Front. Commun.* 3:40.  
doi: 10.3389/fcomm.2018.00040

Bilingual children experience a rapid shift in language preference and input dominance from L1 to L2 upon entering kindergarten when regular contact with L2 starts. Though this change in dominance affects further L1 development, little is known about how various factors shape this. The present study examines the combined influence of different background factors including not only chronological age, age of onset of L2 (L2 AoO), and gender, but also various L1 input measures on L1 receptive and expressive lexical and morphological (case and verb inflections) development in Russian-German bilingual children. For lexical skills, we found a general strong impact of chronological age, gender, and input factors but a differential impact of L2 AoO. Only expressive lexical skills were influenced by language dominance. Morphological development was influenced in the following way: chronological age and gender were most relevant for the acquisition of verb inflection, whereas age, L1 use in the nuclear family and L2 AoO affected the acquisition of case on nouns. This pattern explains the findings of the second series of analyses of longitudinal data, which showed that case is more vulnerable than verb inflection to language attrition—or, taking another perspective—to heritage Russian grammar restructuring.

**Keywords:** input, dominance, home/heritage-language, lexicon, morphology, verbs, nouns, Russian-German

## INTRODUCTION

In migrant families where two parents speak the same home language (henceforth, L1), toddlers and pre-kindergarten children experience dominance in this L1 as opposed to the environment language, i.e., language of the country in which they live (henceforth, L2). Upon entering the regular educational unit, be it nursery school or kindergarten, the input situation changes critically. L2 input begins to dominate and L1 input as well as continued L1 language use radically decreases (Rothman, 2009). According to recent studies (Kohnert and Bates, 2002; Oller and Eilers, 2002; Oller et al., 2011 among others), a shift to L2 preference as a result of this strongly increasing language input takes place within 2–3 years of L2 exposure. This shift in language preference is shaped by various factors and strongly impacts the development of dual languages of bilingual children. While a shift to L2 (input) dominance does not mean achievement of high language proficiency in this language, it might, together with other background factors, substantially impact L1 development. Development in this language may slow down compared with monolingual children's pace of language acquisition (e.g., Flores et al., 2017) or certain grammatical phenomena

may be acquired differently or undergo attrition in certain morphological domains (e.g., Gagarina, 2017). However, the described developmental patterns of home (alternately called heritage, s. below) languages are highly specific for individual language phenomena, shaped by language-specific properties and various acquisition contexts. In other words, this home/heritage language, i.e., a minority language acquired in a migration context, is defined as becoming a non-dominant language once regular contact with the L2 starts (Rothman, 2009; Kupisch and Rothman, 2016). Generally, heritage languages are defined as “languages spoken by the children of immigrants or those who immigrated to a country when young” (Cho et al., 2004, p. 23) and also include—according to Wiley (2005)—languages of the indigenous population, e.g., the Chukchi language in northeastern Russia, as well as earlier colonial languages, e.g., Dutch in South Africa (languages of the indigenous population are not dealt with in our study). Essential components for the definition of heritage languages is the order and degree of their acquisition, being “first in the order of acquisition but not yet acquired because of the individual’s switch to another dominant language” (Polinsky and Kagan, 2007, 369f). Due to switching to the language of the environment, heritage speakers experience various changes in their L1 grammars. These changes follow certain patterns and/or rules, which results in a similarity in heritage grammars, like loss of the pro-drop parameter (Polinsky, 2016) or reluctance to reject ungrammatical or infelicitous material (Bayram et al., 2017), etc. In our study, we examine various patterns in the acquisition of lexicon and of some morphological categories in heritage Russian spoken by second-generation migrant children in Germany within the context of the input dominance shift.

## Impact of L1 Input and Background Factors on L1 Development

Despite changes in input dominance, children continue to receive L1 input to varying degrees at home and during after-school activities. The degree of input as well as the quality of the input is shaped by various factors, such as socio-economic background or presence of bilingual educational institutions among other things. L1 input is usually reduced and significantly determines the development of this language. This has been shown by several studies exploring the influence of input parameters on L1 development in ratings of general language skills (De Houwer, 2007) and lexical (Pearson et al., 1997; Klassert and Gagarina, 2010; Armon-Lotem et al., 2011; Hoff et al., 2012) or grammatical skills (Gutiérrez-Clellen and Kreiter, 2003; Gathercole and Thomas, 2009; Klassert and Gagarina, 2010; Armon-Lotem et al., 2011; Hoff et al., 2012; Thomas et al., 2014; Flores and García, 2017; Rodina and Westergaard, 2017).

The studies use a variety of L1 input measures, such as proportion of language use at home as estimated by the parents (Pearson et al., 1997; Gutiérrez-Clellen and Kreiter, 2003; Hoff et al., 2012), home language situation, i.e., whether one or both parents are L1 speakers (Thomas et al., 2014) or whether the parents are 1st or 2nd generation immigrants (Flores et al., 2017), language policy at home, i.e., whether the parents use

one or both languages at home (De Houwer, 2007; Gathercole and Thomas, 2009; Klassert and Gagarina, 2010; Armon-Lotem et al., 2011; Rodina and Westergaard, 2017 with specification of the proportion, if both languages are used). All these L1 input measures, assessed via questionnaires, revealed significant effects on the different language skills assessed in a variety of L1s and age ranges, even though there is evidence, that the reliability of these parental ratings is low (Carroll, 2017; Marchman et al., 2017).

Some of the studies included additional input measures, in order to account not only for L1 input at home but also for the overall language situation of the children (e.g., Gutiérrez-Clellen and Kreiter, 2003; Rodina and Westergaard, 2017). Gutiérrez-Clellen and Kreiter (2003) assessed the proportion of language input at home (estimated by the parents) in addition to the proportion of language input at school (estimated by the teachers), as well as the number of hours spent reading and doing literacy activities in the target languages of 57 school-children (aged 7;3–8;8) from Latin American migrant families in the United States. An influence on the score of grammatical utterances in an L1 Spanish story-telling task was only found for language input at home. The authors cannot clearly conclude, that L1 in school significantly impacts language development. They suggest, however, that the teachers may have had no objective and independent information about the amount of Spanish input in school and therefore teachers’ ratings may have been less reliable. The same holds for reading activities: these parental ratings could have been guided by social desirability and therefore not be reliable. Rodina and Westergaard (2017) collected detailed cumulative data on language use inside and outside the home with the *Bilingual Language Exposure Calculator* (BiLEC, Unsworth, 2013), which calculates the child’s exposure to the target language in the current year (present exposure) and the total lifetime exposure (cumulative length of exposure) based on parent’s ratings on a Likert scale. Only for the latter measure did a regression analysis reveal an influence on gender marking skills in L1 Russian of Russian-Norwegian children, aged 4;1–7;11. As described above, in another step of their analyses, they included the more global variable of language situation at home (both parents are L1 speakers, or one parent is L1 speaker), which also revealed significantly better L1 gender marking skills for children with two L1-speaking parents. In sum, these studies suggest that input measures from outside the home do not lead to such consistent effects as do parental input measures (see also Gagarina et al., 2014 for combined measure of inside and outside home; Lein et al., 2017). It remains open for discussion whether this is a matter of information reliability or whether the impact of present language use outside the family is not that important for L1 language skills.

While input is a very important factor in language development, it is not the only one. Among the other factors influencing L1 acquisition in bilingual children, the impact of age of onset of L2 (L2 AoO) and length of exposure to L2 (LoE) has been addressed. It is undeniable that these factors are crucial for L2 acquisition, since L2 AoO shape developmental rate and outcomes (although L2 AoO effects on L2 acquisition are mixed) and LoE is the most crucial factor in long-term

input quantity of L2 (e.g., Chondrogianni and Marinis, 2011; Paradis, 2011; Unsworth et al., 2014). In the context of heritage L1 acquisition, L2 AoO can be treated as a measure of input language dominance. This is because it has been shown, that soon after L2 contact begins, a shift toward dominance to L2 also begins (Oller et al., 2011). Moreover, L1 has more time to develop independently if the L2 AoO is later, therefore making the LoE of L2 shorter (Kupisch and Rothman, 2016). Studies examining the influence of L2 AoO/LoE on L1 acquisition have not yield consistent results. Whereas Armon-Lotem et al. (2011) and Gagarina et al. (2014) found no correlation between L2 AoO/LoE and L1 lexical and grammatical skills in two relatively large cohorts of Russian-German and Russian-Hebrew children (Armon-Lotem et al., 2011;  $n = 143$ , age 4–6 years; Gagarina et al., 2014;  $n = 196$ , age 4–7 years), correlations were reported by Lein et al. (2017) in a sample of 14 bilingual children (L1 Portuguese, age 4;9–8;7) in Germany. Receptive and productive lexical tasks and a productive morphosyntactic task (sentence repetition) showed significant correlation with L2 AoO and LoE, but not with a receptive morphosyntactic task (sentence comprehension). Furthermore, Schwartz and Minkov (2014), based on a descriptive comparison of successive and sequential Russian-Hebrew bilingual children ( $n = 9$ ; age 3–4), reported that a low L2 AoO is associated with higher error rates in case marking in spontaneous speech data. Janssen et al. (2015) found that L2 AoO (under consideration of LoE and language spoken at home in a regression analysis) was the only significant predictor for correctness in the case processing task in L1 Russian ( $n = 36$ , age 5;9–7;7 months). Since L2 AoO and LoE should be inter-correlated with the cumulative exposure to a language and to the language situation at home, in order to obtain a more detailed picture it is crucial to explore its influence on L1 language skills together with other input factors in regression analyses.

Another ever-present factor in language acquisition is chronological age. The increase of language skills with chronological age in monolingual language acquisition is a common fact. With increasing chronological age, the total amount of experience with a certain language, present since birth, grows. Cognitive skills mature and language develops. This is not self-evident for the acquisition of the L1 as a minority language in bilingual children. Studies exploring age effects in combination with input factors yield a mixed picture. In two studies on L1 Welsh, increasing L1 abilities with age were reported (analyses with ANOVAs): Thomas et al. (2014) found increasing skills with age in plural marking in a sample of 88 children (age 7–11 years). Interactions with input were not reported. Gathercole and Thomas (2009) reported an effect of age and an interaction with home language as an input measure in a lexical task ( $n = 610$ , age 7–11) and also in a grammatical comprehension task without interaction with input measures ( $n = 248$ , three age groups: 5, 7, 9). For L1 Portuguese in Germany, Flores et al. (2017) also found an influence of age under consideration of the parental input variable (1st or 2nd generation immigrants) and the presence of older siblings, in a sample of 50 subjects (age 6–16). Contrary to this, Lein et al. (2017) found no correlation of age with a variety of language skills for the same population in a younger sample (4;9–8;7).

Finally, we briefly summarize the results of the influence of age in L1 Russian (further details for Russian will be given in the next section). In the studies of Armon-Lotem et al. (2011) and Gagarina et al. (2014), L1 skills correlated with age only for the children living in Germany, but not for the children living in Israel. In Rodina and Westergaard (2017) study on L1 Russian in Norway, age was not a significant predictor for gender-marking skills. It was only the cumulative input measure. It appears that the influence of age on L1 development strongly depends on a conglomerate of various factors, such as the acquisition context and input, and L1 and L2 language-specific properties. The status of a minority language in the society, the institutional support in kindergarten or school, the presence of the language in the environment of a child and, last but not least, the language use patterns at home, and the family language policy in general appeared to determine the speed and success of L1 acquisition more obviously than age.

The influence of gender on L1 skills has so far received much less attention in comparison to various other background factors. In monolingual language acquisition, this influence has been found to occur from the earliest stages of the life span. In early monolingual acquisition, girls exhibit larger vocabularies and produce longer and more complex sentences than boys of the same age (for an overview see Bornstein et al., 2004; Eriksson et al., 2012). Later in L2 acquisition, female learners also outperform male learners in speaking skills (Van Der Slik et al., 2015). Additionally, bilingual girls at age three to eight were shown to be better in the development of their narrative skills in L1 Turkish as compared to the age matched boys (Mavi et al., 2016). Explanations for these findings involve biological, psychological and social explanations (Maccoby, 1966; Bornstein et al., 2004; Eriksson et al., 2012). Recent results of a cross-linguistic and cross-cultural stable gender effect (Eriksson et al., 2012; Van Der Slik et al., 2015) are in favor of a strong biological influence.

All in all, there is still little research on the combined influence of background factors on bilingual L1 acquisition, which is fragile as compared to monolingual acquisition and is more dependent on these background factors. Out of all factors influencing L1 development, L1 input was shown to be the most crucial. But it's differential influence on lexical and grammatical development in interaction with chronological age, L2 AoO, and gender is far from being clear and has, to the best of our knowledge, never been analyzed in combination with other factors (Armon-Lotem et al., 2011; Gagarina et al., 2014; Lein et al., 2017 the studies with the highest number of background factors report only correlations). Moreover, only a few studies examined the influence of input in- and outside the home (e.g., Gutiérrez-Clellen and Kreiter, 2003; Rodina and Westergaard, 2017). Most of the studies on input used either only measures of parental input (e.g., Pearson et al., 1997; De Houwer, 2007; Klassert and Gagarina, 2010; Armon-Lotem et al., 2011) or combined measures for all L1 speakers in the environment of the child (e.g., Gagarina et al., 2014; Lein et al., 2017). Therefore, it is still an open question of whose input most contributes to the heritage language skills of bilingual children. Finally, evidence on the influence of the combination of various background factors on concrete grammatical phenomena

in Russian is sparse. The studies considered either just a limited range of variables (Schwartz and Minkov, 2014; Janssen et al., 2015; Rodina and Westergaard, 2017) or used only combined grammatical measures (Armon-Lotem et al., 2011; Gagarina et al., 2014). A detailed examination might be crucial for understanding patterns of L1 development, because, as shown in the following section, various grammatical phenomena are differentially influenced by the bilingual acquisition context.

## Bilingual Acquisition of Russian as a Heritage Language

The Russian-speaking population is widely spread across the world, with about 30 million people living outside of Russia and the republics of the former Soviet Union (according to the Ministry of Foreign Affairs of the Russian Federation). Depending on the country of residence, history of emigration and various environmental factors, these people have different resources and motivation for maintaining the Russian language and for transferring it to their children—thus, the diversity (the various levels of Russian proficiency and Russian “grammars”) is large. Furthermore, research on the acquisition of Russian as a heritage/home language differs strongly between countries. Here, we will concentrate only on studies dealing with Russian-German bilingualism. This is due to several factors: first, apart from the perceptive and productive lexicon, we are interested in productive morphology. German, as a societal language with poor and non-transparent morphology, impacts the acquisition of heritage Russian in a specific way (Brehmer, 2007; Anstatt, 2008; Dierker, 2009 among others). The second reason has to do with the peculiarities of Russian-speaking diaspora to Germany and its broad network, own print media, local radio- and TV broadcasts, educational offers, doctors, and shops (Soultanian et al., 2008). Russian of the speakers, who emigrated to Germany as adults, but also heritage Russian, which was acquired by children born in Germany, is in fact spoken throughout the country, allowing for its active use in every-day life as well as its stable transfer to further generations [more on Russian(-speaking diaspora) in Germany, see Brehmer, 2007; Anstatt, 2008; Gagarina et al., 2014; Gagarina, 2017].

We now present the results relevant to our study on the acquisition of Russian as a heritage language in the German context, with a focus on lexical skills and morphological categories—verb inflection and case. In bilingual L1 lexicon acquisition, an increase in L1 with chronological age in expressive lexicon was found in cross-sectional studies (Armon-Lotem et al., 2011; Klassert, 2011; Gagarina et al., 2014; Klassert et al., 2014). For productive lexicon, picture-naming of objects and actions, Klassert et al. (2014) showed a bilingual disadvantage (i.e., smaller lexicons in comparison with monolingual peers) that increased with age in comparison to monolingual children. They also found that verb learning was more stable in heritage Russian than noun learning and that the dominance of L2 noun production appeared 3 years after L2 AoO.

Noun and verb morphology was shown to follow various acquisitional patterns in baseline child Russian and in heritage Russian. Noun and verb morphology in Russian is characterized

by richness of inflection, but differs in respect to its syncretism and transparency. Inflection on verbs in Russian is rather homogenous: the majority of verbs build aspectual pairs, both members of which are marked in the past by number—plural and singular, and in singular by the genders—feminine, masculine, and neuter. Imperfective verbs in the present and perfective verbs in the future exhibit synthetic person-number inflection, which is characterized by the one *form-one function* relation (cf. Slobin, 2001). This relation facilitates acquisition of this inflection and leads to low overgeneralization rates. Case is generally considered to be “one of the most heterogeneous nominal morphological categories” (Eisenbeiss et al., 2009, p. 369) in the languages of the world, and Russian is not an exception. Nouns in the three declension classes exhibit six cases (in singular and plural) with the same marking for various cases within one declension class and across singular and plural number, e.g., *myši* “mouse-GEN.SG or -DAT.SG or -NOM.PL.” Additionally, Russian exhibits differential object marking, with animate objects having similar inflections to the genitive case and inanimate objects to the nominative case, e.g., *myšej* “mouse-GEN or ACC singular” and *stul* “chair-NOM.SG or -ACC.SG.” The flexibility of stress on nouns and vowel reduction in the unstressed final inflections increases syncretism and opacity of the case system. This difference in the target noun and verb morphology leads to differences in timing and path of acquisition.

Studies on the acquisition of verb inflection showed its early and stable development, especially for person marking [however, past tense marking of verbs, which includes agreement with the subject in number and gender was found to be vulnerable in eight Russian-Hebrew bilingual children aged 3;6–5;0 (Gagarina et al., 2007)]. Early acquisition of person-marking on verbs in heritage Russian was shown to be similar to that of monolinguals in one case study: Gagarina (2008) used a longitudinal corpus to establish the stages in L1 bilingual development of verb categories; she found similarities in timing and acquisition path of verb inflection between a simultaneous bilingual child and four monolingual children (cf. Kiebzak-Mandera, 2000). Generally, bilingual children were shown to master person marking on verbs in L1 similarly to monolingual children, with high acquisition speed (Xanthos et al., 2011) and low error rate, as opposite to past tense marking.

Acquisition of case-marking follows another pattern. For monolingual acquisition of Russian, Gagarina and Voeikova (2009) performed a multiple-case longitudinal study with four children and reported a very early emergence of all case oppositions (in singular), but not its productive use, which was shown to fully develop by age three, i.e., later than tense-person inflection on verbs. They suggested that the acquisition of case inflection corresponds to inflectional classes of nouns and is driven by transparency and iconicity of form-function meaning. Furthermore, they distinguished between several degrees of productiveness dealing with the variability in use of a given case form with various inflectional classes and with its frequency. They concluded that “various types of morphophonemic markings and the contrastive forms construct a system of cases that approaches the target Russian” (Gagarina and Voeikova, 2009, p. 212). Thus, they explained the acquisition



of the case marking system via the transparency, non-syncretism and frequency of its single elements. For example, both in monolingual and bilingual acquisition of accusative marking in the first declension inflectional class of *-a*-nouns, class is the first of the six cases in Russian to be acquired. Whereas in monolingual children it usually remains stable and does not undergo any changes, bilingual children might modify their use of this inflection. In a longitudinal study, Gagarina (2011) documented the productive use of accusative inflection *-u* (the first inflectional class of *-a*-nouns) in a simultaneous bilingual child at age three: *klouna* “clown-ACC,” but the loss of this inflection, i.e., *\*kloun* at age six. This was despite L1 language input remaining stable and the child receiving school education in both languages. Such a loss of already acquired categories or constructions affects only those areas of the grammatical system that are less transparent and characterized by high syncretism, low frequency and later age of acquisition (cf. Gagarina and Reichel, 2013). This pattern is a clear example of attrition of an inflectional category, or of marking of accusative case within one declension class; this process of attrition is traceable only if longitudinal observations are made and the correct productive use in the earlier acquisition phase has been observed.

While these studies suggest that language-specific morphological properties and children’s own preferences may lead to case errors, recent research shows that (i) the percentage of overgeneralisations in children’s speech is comparatively low, and (ii) case errors mostly occur during a limited period of time. In particular, children acquiring Slavic and Baltic languages, such as Croatian, Lithuanian, Russian, generally exhibit low overgeneralisation rates (Voeikova and Gagarina, 2002; Katičić, 2003, p. 110–112; Savickiene, 2002, p. 131–133). The authors view these findings as evidence for a strong reliance on rote-learned forms and transparent analogies in early grammatical development.

Generally, Gagarina (2017) specified that for the acquisition of heritage Russian in Russian-German bilingual children, as language acquisition progresses, some already acquired elements, which are non-transparent, not the first to be acquired, and “stabilized” in their productive use, undergo attrition. As a result of this process, the restructuring of (the elements of) the grammatical system takes place.

## THE STUDY: GOALS AND RESEARCH QUESTIONS

All in all, bilingual L1 acquisition was shown to be fragile as compared to monolingual acquisition and more dependent on various background factors, contributing to the non-dominance of L1 in daily life (Oller et al., 2011). Our study aims to close several gaps in research on heritage L1 Russian acquisition in bilingual children. We address not only productive and perceptive lexicon, but analyze productive L1 morphology, in particular the production of person-number inflection on verbs in present tense and case on nouns.

In the first series of analyses we inspect the impact of various background factors on L1 skills in a more differentiated way than

previous studies, which used either combined input measures for family and friends (Gutiérrez-Clellen and Kreiter, 2003; Armon-Lotem et al., 2011; Gagarina et al., 2014) or only parental input measures (De Houwer, 2007; Pearson, 2007). We distinguish between L1 use with nuclear family members, i.e., parents and siblings, and L1 use with other people, i.e., other members of the family or friends, and investigate its influence as combined with chronological age, L2 AoO and gender on L1 performance in the selected domains. This differentiated view allows for a more fine-grained examination of the effects of L1 input on language performance in specific language domains. Furthermore, these analyses will show which of the child’s communication partners’ L1 use mostly impacts L1 development in contexts in which L1 input loses its dominance.

In the second series of analyses, we trace the longitudinal development of L1 lexicon and morphology (person-number inflection on verbs and case on nouns) in two cohorts of preschool children. We thereby extend the results of the first series of analyses by exploring the age factor in more detail. This will allow us to provide a deeper insight into the developmental patterns of these particular domains in a non-dominant context of L1 acquisition. Previous studies suggest, that the acquisition of noun and verb morphology in L1 Russian, in particular case and person-number inflection, is affected differently by this acquisition context (Gagarina, 2008; Gagarina and Reichel, 2013) and that lexical abilities increase with age in Germany but not in other countries (Armon-Lotem et al., 2011 for Israel). However, no studies so far have investigated the longitudinal L1 development of Russian in a larger sample in these different domains.

Our study explores selected language domains of Russian-German bilingual children and addresses the following research questions: (1) How do the background factors chronological age, gender, L2 AoO and L1 input (differentiated across the nuclear family and other people) impact L1 acquisition of receptive and productive lexicon and of two domains of morphology—accusative and dative case on nouns, 1st and 2nd person on present tense verbs? (2) Which developmental changes in L1 lexicon and morphology (case on nouns and person on verbs) are observed between the 3rd and 4th and between the 4th and 5th years of age in these children?

## METHOD

### Participants

The data come from a large sample of Russian-German bilingual children. Most of the data were gathered at Leibniz-ZAS Berlin from 2008 to 2017 for *Russian language proficiency test for multilingual children* (Gagarina et al., 2010, 2015) and in the context of the Berliner Interdisciplinary Association for Multilingualism (BIVEM) Project. Another small set of the data comes from the lab of Prof. Cornelia Hamann in northern Germany (Bremen/ Oldenburg). All participants showed, according to the parental questionnaires and teachers, no motoric, cognitive, socio-psychological or any other disorders. For our analyses, we used different subsets of the data, which are described in the respective sections in the results.

The results of *Analysis 1, exploring the impact of background factors on L1 development*, are based on a subsample of 213 Russian-German bilingual children between 26 and 98 months ( $M = 52.76$ ,  $SD = 17.41$ ), 44.6 female. 91.5% of the data were gathered in Berlin. The data of 18 children come from Bremen/Oldenburg. For all children, one or more parts of the language measures were assessed, and the questionnaire was filled out by the parents. Due to data collection problems, there are missing data for all measures (see **Table 1**). The evaluation of the questionnaire (described in section Background Measures) concerning L2 AoO, revealed that 14.6% ( $n = 31$ ) of the sample came in regular contact with L2 German below 18 months of age, 42.3% ( $n = 90$ ) between 18 months and 3;05 years and 11.7% ( $n = 25$ ) between 3;06 and 5;05 years. For 31.5% ( $n = 67$ ) the parents did not indicate L2 AoO for their child. The descriptive statistics for input and language measures are presented in **Table 1**.

**Table 2** depicts the number of people specified in the questionnaire for L1 use with the child. In 95.8% of the cases, both parents' L1 use was specified, and in 62.9% that of one or more siblings. As described in section Background Measures, these data were combined for L1 use in the nuclear family. The L1 use of other people was specified for 85.9% of the children in the sample. The first two persons were always grandparents. Additional other people were also grandparents but also other relatives (cousins, aunts and uncles), and further persons close to the child (friends and neighbors).

*Analysis 2, exploring the longitudinal development of L1 lexicon and morphology*, are based on a subsample of 116 three- and four-year-old Russian-German bilingual children living in Berlin. The L1 language skills of these children were tested twice with an interval of ~1-year (interval between testing 1 (T1) and testing 2 (T2)  $M = 11.65$  months,  $SD = 0.71$ ,  $range = 10$ –14 months). The 3-year-old sample (AG3) comprised 58 children (48.3% female). Their mean age at T1 was 42.02 months ( $SD = 3.40$ ,  $range =$

36–47 months) and at T2 53.55 months ( $SD = 3.78$ ,  $range = 47$ –61). The 4-year-old sample (AG4) consisted of 58 children (39.7% female; age T1  $M = 52.05$  months,  $SD = 3.17$ ,  $range = 48$ –59; age T2  $M = 63.81$  months,  $SD = 3.18$ ,  $range = 59$ –71). As presented in **Table 5**, not all children completed all language subtests. For case and verb inflection especially, the samples are smaller.

## Linguistic Measures: Productive and Receptive Lexicon

For the present study, several subtests from the *Russian Language Proficiency Test* (SRUK, Gagarina et al., 2010, 2015) were used, namely, receptive and productive lexicon and productive morphology: case on nouns and person-number on verbs. The productive lexicon was tested by means of a picture-naming task, consisting of nouns and verbs (for each word category there were two training items and 26 test items). The children were shown individual pictures and asked, “What is this?” or “What is s/he doing?” The test items were chosen based on unambiguous identifiability of the pictures, frequency of the item and semantic field (for verbs, the category of aspect was considered as well). For some items, several responses were accepted as correct. The comprehension of individual words was tested by means of a picture-selection task for nouns and verbs (for each word category there were one training item and 10 test items). The tester presented the word auditorily and the child had to choose the correct picture from a group of four pictures by pointing to it. The three distractors were composed of a semantically-related, a phonologically-related, and an unrelated item of the same part of speech. Again, the test items belonged to different frequency ranges and were controlled for unambiguous identifiability with monolingual adults and children.

## Linguistic Measures: Productive Morphology

Two linguistic subtests testing case on nouns and person-number on verbs were used in order to examine productive morphology. The case subtest consisted of two training questions and six elicitation questions. Three of the questions elicit the accusative and three the dative case. These two cases were chosen because they both oblique cases and are central for the Russian case system, show stable use in monolingual children by age three (Gagarina and Voeikova, 2009) and exhibit, in some context, direct correspondences to accusative and dative cases in German, e.g., *Papa darit devočke-DAT cvety-AKK* (Russian), *Der Vater schenkt dem Mädchen-DAT die Blumen-AKK* (German) “The father presents the flowers to the girl.” The case subtest has two parts: in the introduction, the child is familiarized with the circus picture, accompanied by a story of a circus where various characters, which are presented as pairs of two puzzle pieces, are friends. The tester names all items in the nominative case: *Here is a lion, an elephant, a monkey, a snake*, etc. Then the child answers the elicitation question and puts pieces of two-part puzzles together. The elicitation question for dative is *Komu nravitsja X? (Who (-DAT) does the lion like?)*, and for accusative *Kogo iščet X? (Whom (-ACC) X is looking for?)*.

**TABLE 1 |** Descriptive statistics for input measures and language measures.

	<i>N</i>	Min	Max	<i>M</i>	<i>SD</i>
L1 use nuclear family	213	0.50	4.00	2.64	0.88
L1 use other	183	0.00	4.00	2.68	1.13
Receptive lexicon	206	2	20	14.52	4.11
Productive lexicon	204	0	51	22.23	13.22
Case	153	0	6	2.29	2.22
Verb inflection	118	0	12	9.65	3.49

**TABLE 2 |** Number of people [*N* (%)] specified in the questionnaire for language use with the child.

	0	1	2	3	4	5	6
Parents		9 (4.2)	204 (95.8)				
Siblings	79 (37.1)	96 (45.1)	28 (13.1)	8 (3.8)	2 (0.9)		
Other	30 (14.1)	45 (21.1)	61 (28.6)	43 (20.2)	29 (13.6)	3 (1.4)	2 (0.9)

Verb inflection is tested for the first and second-person singular imperfective present. Although the first and second-person singular imperfective present is acquired after the third person, all tense-person inflections occur early and are used in a target-like fashion prior to age three (Gvozdev, 1949; Gagarina, 2008). The test consists of 2 training items and 6 test items. The child and tester perform certain actions and the child is asked *Who is doing what?* S/he has to name the action, e.g., *Ja igraju* “I’m playing” or *Ty čitaješ* “You’re reading.”

The number of correct responses (according to Gagarina et al., 2015) for each subtest was used as the child’s final score in the data analysis.

All data were collected in monolingual modus in a separate room in the kindergartens or schools by a native speaker after the parental consent forms were signed. The parents and teachers were informed about the goals, content and procedure of the studies.

## Background Measures

A questionnaire was administered for the gathering of detailed information on each child’s individual language acquisition context and input situation. The entire questionnaire is published in Gagarina et al. (2010). For the present study, the following two parts were used: In the questionnaire, the parents had to indicate how old the child was when it came into regular contact with German, i.e., L2 AoO as one of three categories: 1. below 18 months, 2. between 18 months and 3;05 years, 3. between 3;06 and 5;05 years. In another part of the questionnaire, the parents were asked to rate the child’s language use with his/her mother, father, siblings and other people who were in frequent contact with the child: person X speaks (1) only German, (2) little Russian, much German, (3) Russian and German equally, (4) much Russian, little German, (5) only Russian. For the data analysis, this was converted into a 5-step scale according to the numbers given before [from 0 (only German) to 4 (only Russian)]. For each child, the mean language use of L1 Russian was calculated separately for the nuclear family (*L1 use nuclear family* = mean of language use for parents and siblings) and for other persons specified by the parents (*L1 use other*).

The parents filled out the questionnaire at home, without the guidance of an instructor.

## Data Analysis

The data were analyzed using SPSS statistics 24.

To evaluate the impact of background factors on L1 development (Analysis 1), in the first step we calculated the correlations between language measures and background factors, using pairwise deletion of missing cases. Correlations were assessed by calculating  $r$  with Pearson correlation between metrical variables (language measures and input measures) and with Spearman’s rank correlation between metrical and the ordinal variable L2 AoO (interpretation according to Cohen, 1988:  $r = 0.10$ – $0.29$  small,  $r = 0.30$ – $0.49$  medium,  $r = 0.50$ – $1.0$  large). Correlations between the nominal variable age and the metrical/ ordinal variables were assessed with the eta correlation ratio (interpretation according to Cohen, 1988:  $\eta = 0.01$ – $0.03$  small;  $\eta = 0.04$ – $0.15$  medium;  $\eta > 0.16$  high). In the second step, we performed multiple regression models for each language

competence measure as a dependent variable (sum for individual child) with listwise deletion of missing cases. All factors were chosen on the basis of the results of the initially performed correlations, and were entered simultaneously as predictors.

To explore the longitudinal development of L1 lexicon and morphology (Analysis 2), repeated measures ANOVAs were performed to explore the influence of testing time and age on the different language skills. Age (with the two age groups AG3 and AG4) served as between-subject factors, and *sum correct* of the certain language measure as a within-subject factor (with testing time T1 and T2 as levels). *Post-hoc t*-tests were performed in case of significant between-subject effects and significant interactions. Additionally, effect sizes are reported, partial  $\eta^2$  for ANOVAs and  $d$  for *t*-tests. In the classification of effect sizes, we follow Cohen (1988) with  $d = 0.2$  small,  $d = 0.5$  medium,  $d = 0.8$  large, and Döring and Bortz (2006) with *partial*  $\eta^2 = 0.001$  small, *partial*  $\eta^2 = 0.10$  medium, *partial*  $\eta^2 = 0.25$  large.

The sum of included cases is reported as  $n$  for correlations and as  $df$  for regressions and ANOVAs.

## RESULTS

### Analyses 1: The Impact of Background Factors on L1 Development

#### Correlations Between Language Measures and Background Factors

First, we assessed correlations between the different language measures and the background factors. **Table 3** shows the correlations between the language measures and all background factors except gender (the correlations between this binary variable and the other variables are presented afterward).

All language measures are significantly correlated between each other. The correlations between *receptive lexicon* and *expressive lexicon* [ $r_{(200)} = 0.715, p < 0.001$ ], *receptive lexicon* and *verbal inflection* [ $r_{(111)} = 0.542, p < 0.001$ ] and between *expressive lexicon* and *case* [ $r_{(143)} = 0.734, p < 0.001$ ] are high. Correlations between *receptive lexicon* and *case* [ $r_{(146)} = 0.461, p < 0.001$ ], *expressive lexicon* and *verbal inflection* [ $r_{(110)} = 0.494, p < 0.001$ ] and between *case* and *verbal inflection* [ $r_{(116)} = 0.374, p < 0.001$ ] are moderate.

The language measures are significantly correlated with the background factors age [*receptive lexicon*  $r_{(204)} = 0.599, p < 0.001$ , *expressive lexicon*  $r_{(202)} = 0.607, p < 0.001$ , *case*  $r_{(152)} = 0.328, p < 0.001$ , *verbal inflection*  $r_{(117)} = 0.231, p = 0.011$ ] and *L1 use in the nuclear family* [*receptive lexicon*  $r_{(204)} = 0.191, p = 0.006$ , *expressive lexicon*  $r_{(202)} = 0.385, p < 0.001$ , *case*  $r_{(152)} = 0.316, p < 0.001$ , *verbal inflection*  $r_{(117)} = 0.182, p = 0.048$ ] to different degrees ranging from small to large. L2 AoO is significantly correlated with all language measures [small to moderate correlations, *receptive lexicon*  $r_{S(138)} = 0.324, p < 0.001$ , *expressive lexicon*  $r_{S(140)} = 0.449, p < 0.001$ , *case*  $r_{S(99)} = 0.280, p = 0.005$ ] except *verbal inflection* [ $r_{S(75)} = 0.181, p = 0.114$ ]. *L1 use other* shows no significant correlations to the language measures. Furthermore, there are intercorrelations between the background factors: *L1 use in the nuclear family* and *L1 use other* correlate moderately [ $r_{(181)} = 0.407, p < 0.001$ ]. Correlations with small effect sizes appear between age and *L1 use other* [ $r_{(181)} =$

**TABLE 3 |** Correlations between language measures and background factors with Pearson correlation, <sup>a</sup>Spearman's rank correlation, \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

	2.	3.	4.	5.	6.	7.	8. <sup>a</sup>
1. Receptive lexicon	0.715***	0.461***	0.542***	0.599***	0.191**	−0.018	0.324***
2. Expressive lexicon	1	0.734***	0.494***	0.607***	0.385***	0.036	0.449***
3. Case		1	0.374***	0.328***	0.316***	−0.050	0.280**
4. Verbal inflection			1	0.231*	0.182*	0.001	0.181
5. Age				1	−0.064	−0.267***	0.226*
6. L1 use nuclear family					1	0.407***	0.229***
7. L1 use other						1	0.159
8. L2 AoO							1

N's range from 77 to 206 due to missing data (for details see text).

−0.267,  $p < 0.001$ ], *age* and *L2 AoO* [ $r_{S(144)} = 0.226$ ,  $p < 0.001$ ] and *L2 AoO* and *L1 use nuclear family* [ $r_{S(144)} = 0.229$ ,  $p < 0.001$ ]. *Age* and *L1 use nuclear family* [ $r_{(211)} = -0.064$ ,  $p = 0.353$ ] as well as *L1 use other* and *L2 AoO* [ $r_{S(132)} = 0.159$ ,  $p = 0.066$ ] are not significantly correlated.

The correlations of *gender* with the other variables were assessed with eta correlation ratio because *gender* is a binary variable. The analysis revealed that, concerning the language measures, *gender* is moderately correlated with *receptive lexicon* ( $\eta = 0.149$ ,  $n = 198$ ) and *verbal inflection* ( $\eta = 0.134$ ,  $n = 110$ ), highly correlated with *expressive lexicon* ( $\eta = 0.250$ ,  $n = 197$ ) and weakly correlated with *case* ( $\eta = 0.010$ ,  $n = 145$ ), all in favor for the girls. Calculating eta correlation ratio of *gender* with the other background factors showed no correlation with *age* ( $\eta = 0.007$ ,  $n = 204$ ), a small correlation with *L2 AoO* ( $\eta = 0.017$ ,  $n = 146$ , being a girl is associated with a later L2 AoO) and medium correlations with *L1 use nuclear family* ( $\eta = 0.137$ ,  $n = 204$ ) and *L1 use other* ( $\eta = 0.050$ ,  $n = 182$ ), both indicating that the L1 use with girls is higher than with boys.

In sum, the correlations show that the language measures are more strongly correlated between each other than with the background variables. Nevertheless, nearly all background variables are correlated to different degrees with at least most of the language measures. The exception is *L1 use other*, which is not significantly correlated to the assessed language skills. Moreover, it is intercorrelated with the other input measure *L1 use nuclear family*. Therefore, to reduce the model complexity and to avoid multicollinearity in the regression, the background variable *L1 use other* will be excluded from the following regression models.

## Regression Analysis

To evaluate the influence of background factors on L1 language competence, we ran multiple regression models for each language competence measure as dependent variable and chronological *age* in months, *gender*, *L1 use nuclear family* and *L2 AoO* as predictors, all entered simultaneously. *L1 use other* was not included as predictor because of the missing correlation to the language measures and the intercorrelation to *L1 use nuclear family*. Detailed results for the predictors of all regressions described in this part are presented in **Table 4**.

For the *receptive lexicon* the regression model with *age*, *gender*, *L1 use nuclear family*, and *L2 AoO* as simultaneous predictors was significant [ $F_{(4,135)} = 35.67$ ,  $p < 0.001$ , *mean VIF* = 1.13].

*Age*, *gender* and *L1 nuclear family* were significant predictors and explained 50% of the variance in the test scores (*adj. R*<sup>2</sup> = 0.50). The strongest standardized beta coefficient was found for *age* ( $\beta = 0.592$ ). For *L1 use nuclear family* ( $\beta = 0.194$ ) the standardized beta coefficient was slightly higher than for *gender* ( $\beta = -0.176$ ). *L2 AoO* was not a significant predictor for receptive lexical skills. In sum, a higher age, an increased amount of Russian language use at home and being a girl lead to better receptive lexicons.

For the *expressive lexicon* the predictors explained 67% of variance in the data [*adj. R*<sup>2</sup> = 0.67;  $F_{(4,137)} = 70.96$ ,  $p < 0.001$ , *mean VIF* = 1.1]. In this model all predictors had significant coefficients (for details see **Table 4**), with the largest standardized beta for *age* ( $\beta = 0.601$ ), followed by *L1 use nuclear family* ( $\beta = 0.281$ ), *L2 AoO* ( $\beta = 0.196$ ), and *gender* ( $\beta = -0.172$ ), in favor of girls. So, the older the children, the more the nuclear family uses Russian, the later their L2 AoO of German and if the children are female, the better the expressive lexicon of the children in our sample.

For *case marking skills* the multiple regression model with the selected predictors described above was again significant [ $F_{(4,96)} = 7.60$ ,  $p < 0.001$ , *mean VIF* = 1.0]. Significant predictors (ordered by impact according to standardized coefficients) were *L1 use nuclear family* ( $\beta = 0.283$ ), *age* ( $\beta = 0.278$ ), and *L2 AoO* ( $\beta = 0.179$ ). They explained 21% of variance in the data (*adj. R*<sup>2</sup> = 0.21). *Gender* was clearly no significant predictor for case marking skills.

Also for *verb inflection* the multiple regression with *age*, *gender*, *L1 use nuclear family* and *L2 AoO* was significant [ $F_{(4,72)} = 4.30$ ,  $p = 0.004$ , *mean VIF* = 1.0]. Only *age* (with the highest standardized  $\beta = 0.388$ ) and *gender* ( $\beta = -0.200$ , again in favor of the girls) were significant predictors and explained 15% of variance in the test scores (*adj. R*<sup>2</sup> = 0.15). *L1 use nuclear family* and *L2 AoO* were clearly not significant.

In sum, the regression analysis revealed an influence of chronological age on all tested areas of L1 Russian in the sample of Russian-German bilingual children. However, the age range in our sample was large (26–98 months), so this pattern was to be expected. Nevertheless, it was very important to control for age, in order to obtain a comprehensive picture of the influence of the other factors of interest. This turned out to be different for each assessed language skill. Moreover, the amount of explained variance for the lexical tasks was clearly larger than for the



**TABLE 4 |** Multiple regression models for each language competence measure.

Receptive lexicon					
Factors	B	SE B	$\beta$	t	p
(constant)	4.989	1.482		3.366	0.001
Age	0.151	0.016	0.592	9.376	<0.001
Gender	-1.468	0.509	-0.176	-2.887	0.005
L1 use nuclear family	0.971	0.324	0.194	2.999	0.003
L2 AoO	0.638	0.454	0.093	1.405	0.162
$R = 0.712$ , $adj. R^2 = 0.499$ , $F_{(4,135)} = 35.67$ , $p < 0.001$					
Expressive lexicon					
(constant)	-14.356	3.733		-3.846	<0.001
Age	0.461	0.039	0.601	11.848	<0.001
Gender	-4.429	1.277	-0.172	-3.467	0.001
L1 use nuclear family	4.357	0.812	0.281	5.364	<0.001
L2 AoO	4.100	1.103	0.196	3.717	<0.001
$R = 0.821$ , $adj. R^2 = 0.665$ , $F_{(4,137)} = 70.96$ , $p < 0.001$					
Case					
(constant)	-4.189	1.306		-3.207	0.002
Age	0.042	0.014	0.278	3.075	0.003
Gender	0.278	0.407	0.061	0.683	0.496
L1 use nuclear family	0.810	0.260	0.283	3.116	0.002
L2 AoO	0.685	0.346	0.179	1.982	0.050
$R = 0.490$ , $adj. R^2 = 0.209$ , $F_{(4,96)} = 7.60$ , $p < 0.001$					
Verb inflection					
(constant)	5.950	2.199		2.706	0.008
Age	0.073	0.020	0.388	3.647	<0.001
Gender	-1.105	0.590	-0.200	-1.873	0.049
L1 use nuclear family	0.267	0.420	0.067	0.635	0.527
L2 AoO	0.265	0.525	0.054	0.504	0.615
$R = 0.439$ , $adj. R^2 = 0.148$ , $F_{(4,72)} = 4.30$ , $p = 0.004$					

morphological tasks, where power was lower because of the moderately smaller sample size due to missing data.

## Analyses 2: Longitudinal Development of L1 Lexicon and Morphology

The following analyses explore the longitudinal development of L1 lexicon and morphology in 3- and 4-year-old children over the course of 1 year. For all language measures, means from T1 to T2 increase for the total sample as well as for the single age groups numerically, as presented in **Table 5**. First, we will take a closer look at the morphological tests. For *case*, which was tested with 6 items, the younger children (AG3) performed correctly for an average of 34% at T1 and correctly for 35.5% at T2, and the older children (AG4) 31.8% at T1 and 45.3% at T2. For *verb inflection*, which was tested with 12 items, the means of correct responses in percent were higher (AG 3: T1  $M = 45.8\%$ , T2  $M = 67.6\%$ ; AG 4: T1  $M = 66.3\%$ , T2  $M = 82\%$ ).

The ANOVAs revealed a significant main effect of *testing time* for all language measures (for a report of these ANOVAs see **Table 5**). The children therefore improved significantly from T1 to T2 in all language skills assessed. For *case*, the effect size was

only small (partial  $\eta^2 = 0.057$ ), for all other language measures there was a medium effect (partial  $\eta^2 > 0.14$ ).

An effect of *age* was found in the ANOVAs only for *receptive lexicon*,  $F_{(1,114)} = 6.01$ ,  $p = 0.016$ , partial  $\eta^2 = 0.050$ , and *verb inflection*,  $F_{(1,63)} = 5.43$ ,  $p = 0.023$ , partial  $\eta^2 = 0.079$ . This indicates, that in these two measures there are differences between the age groups. *Post-hoc* tests with independent *t*-test revealed a better performance of the older age group (AG4) for *receptive lexicon* than the younger age group (AG3) at both testing times [T1:  $t_{(114)} = -2.53$ ,  $p = 0.013$ ,  $d = 0.24$ ; T2:  $t_{(114)} = -2.14$ ,  $p = 0.035$ ,  $d = 0.20$ ], for *verb inflection* only at T1 [ $t_{(63)} = -2.17$ ,  $p = 0.033$ ,  $d = 0.27$ ] but not at T2 [ $t_{(63)} = -1.75$ ,  $p = 0.086$ ,  $d = 0.22$ ]. As mentioned before, there was no significant main effect of *age* for *expressive lexicon* [ $F_{(1,113)} = 3.72$ ,  $p = 0.056$ , partial  $\eta^2 = 0.032$ ] and *case* [ $F_{(1,73)} = 0.23$ ,  $p = 0.634$ , partial  $\eta^2 = 0.003$ ]. This suggests, that older and younger children perform equally well at individual testing times.

None of the ANOVA found an interaction between *testing time* and *age*, indicating that younger children improved similarly in the different language measures to older children [*receptive lexicon*:  $F_{(1,114)} = 0.251$ ,  $p = 0.617$ , partial  $\eta^2 = 0.002$ ; *expressive lexicon*:  $F_{(1,113)} = 0.784$ ,  $p = 0.378$ , partial  $\eta^2 = 0.007$ ; *case*:

**TABLE 5 |** Descriptive statistics for the total sample (all) as well for the single age groups (AG3, AG4) and report of main effects of ANOVAs for testing time (T1-T2).

Language measure	Group	N	T1 [M (SD)]	T2 [M (SD)]	F (df)	P	Partial eta <sup>2</sup>
Receptive lexicon	All	116	13.14 (3.90)	14.11 (4.01)	18.97 (1,114)	<0.001	0.143
	AG3	58	12.24 (3.94)	13.33 (4.12)			
	AG4	58	14.03 (3.69)	14.9 (3.77)			
Expressive lexicon	All	115	17.37 (12.79)	21.22 (13.91)	34.92 (1,113)	<0.001	0.236
	AG3	57	15.35 (11.79)	18.61 (13.71)			
	AG4	58	19.36 (13.51)	23.78 (13.74)			
Case	All	75	1.97 (2.08)	2.47 (2.24)	4.39 (1,73)	0.040	0.057
	AG3	32	2.06 (2.08)	2.13 (2.27)			
	AG4	43	1.91 (2.10)	2.72 (2.22)			
Verb inflection	All	65	6.89 (4.62)	9.09 (4.02)	15.73 (1,63)	<0.001	0.200
	AG3	28	5.5 (4.73)	8.11 (4.42)			
	AG4	37	7.95 (4.30)	9.84 (3.58)			

$F_{(1,73)} = 3.23$ ,  $p = 0.077$ , partial  $\eta^2 = 0.042$ ; *verb inflection*:  $F_{(1,63)} = 0.40$ ,  $p = 0.531$ , partial  $\eta^2 = 0.006$ ].

This longitudinal analysis clearly show that L1 Russian language skills of 3 and 4-year-old children develop within 1 year. For case, this effect is smaller than for verb inflections and lexicon. Although there is an increase, even the oldest children (AG4 at T2) perform at a very low level in the case test and are, as a group, far from the full acquisition of the case target system.

## DISCUSSION

This study aimed at examining the role of various background factors in L1 acquisition in Russian-German bilingual children when L1 input shifts from dominant to non-dominant upon entering an educational unit in Germany. In particular, we aimed at establishing a scrutinized picture of the impact of the background factors chronological age, gender, individual L1 input, and L2 AoO on the development of heritage Russian. We furthermore aimed at tracing the development of lexicon and two morphological categories—case on nouns, and tense-person on verbs—longitudinally in the home language, to explore in detail the impact of the age factor on these domains.

The multiple linear regression analysis of the influence of background factors on different language skills reveal a differentiated picture for individual language domains. At first, chronological age has a strong impact on all tested areas of language acquisition, even under consideration of other background factors in the same statistical model. For all tested domains except case, age was the predictor with the highest coefficient. This shows that receptive and expressive lexical skills as well as case marking and verb inflection increase with chronological age in heritage Russian for children living in Germany. This finding confirms and extends the results of previous studies, which consider age as an isolated factor (e.g., Armon-Lotem et al., 2011; Gagarina et al., 2014; Klassert et al., 2014) and displays the vivid language situation of the Russian diaspora in Germany as the background for L1 development (cf. Soultanian et al., 2008). On the one hand, one could claim that

this strong influence of age is caused by the large age range in our sample (26–98 months). One must keep in mind though, that increasing L1 language abilities are not self-evident, due to the important role of input factors (e.g., Lein et al., 2017; Rodina and Westergaard, 2017). Our longitudinal analyses confirmed the influence of age by revealing significant increases over 1 year in abilities in the respective language domains in 3 to 4-year-old children.

Secondly, we found significant effects of children's gender on all language measures except case, as manifested in an advantage for girls. This is the first study that confirms the gender gap for L1 bilingual acquisition, which was also reported for monolingual (e.g., Bornstein et al., 2004; Eriksson et al., 2012) and L2 acquisition (Van Der Slik et al., 2015). Interestingly, gender was also correlated with the input situation in our data: being a girl is correlated with a later L2 AoO and a higher L1 use with the family and other people. These findings point out the importance to control for gender if the influence of the acquisition context is of interest. This was considered in our study, and the significant contribution of gender to expressive and receptive lexical skills, as well as to verbal inflection skills, demonstrate that its L1 acquisition is driven by this biological feature.

Concerning L1 use in the child's environment, this study differentiated between L1 use in the nuclear family (parents and siblings) and L1 use with other people. This was done with the aim of scrutinizing to what extent the L1 input situation of a child must be assessed to most comprehensibly evaluate its influence on his/her L1 language skills. Our data suggest, that it is obviously more reliable to assess L1 use in the nuclear family, since the L1 use with other people was not correlated with any of our language measures. Parents' answers varied very much concerning the number of people outside the nuclear family whom they reported to have language use with their child (this factor could have been reduced by conducting the questionnaire in the presence of an instructor). Additionally, communication with these partners might be of various frequency and intensity. In any case, the estimate of the language use of other people with their child might be difficult for the parents, since it has been shown, that they are not even fully reliable in estimating

their own language use (Carroll, 2017; Marchman et al., 2017). Despite these problems with the self-ratings of language use, we found that the L1 use in the nuclear family is the second most important predictor for lexical skills, after age. This confirms and extends the results of numerous previous studies on lexical skills, which found correlations between family/parental input and lexical skills (e.g., Pearson et al., 1997; Klassert and Gagarina, 2010; Armon-Lotem et al., 2011; Hoff et al., 2012). For morphological skills in L1 Russian, our results revealed an interesting dissociation: L1 use in the nuclear family was the most important predictor for case marking skills but did not influence verbal inflection skills, suggesting that verbal inflection is robust to input effects, whereas case is especially sensitive to it (we turn to this point later on in the discussion). In sum, one cannot conclude from our results that L1 use outside the family is not important for L1 language development. Rather, it is very hard to assess, and L1 use within the family (and presumably L1 use in general) differentially impacts different morphological phenomena.

An influence of age of onset of L2 appeared in our multiple regression analyses only for expressive lexicon and case. The later the children came into contact with the L2, the better they performed in these L1 domains. Concerning lexical abilities, our results replicated the findings of Lein et al. (2017), who found for Portuguese L1 acquisition, that lexical abilities are correlated with L2 AoO. A new result is, that under consideration of other background factors in the regression model, L2 AoO is a significant predictor only for expressive lexicon. L2 AoO in the context of L1 acquisition reflects language dominance (Oller et al., 2011), in the sense that with increasing duration of L2 contact, the amount and relevance of input as well as the amount of L2 language use increases and becomes prevalent for the child, so that L1 use and relevance decreases. Oller et al. (2011) summarize several studies which have documented a large receptive-expressive gap in L1 lexical abilities, concluding that lexical retrieval is very much affected by the change of dominance (for similar results and argumentation see Yan and Nicoladis, 2009). This is mirrored in our data, in the differential influence of L2 AoO and also on morphological categories. L2 AoO differentially impacts our data. Case inflection in heritage Russian is very sensitive to L2 AoO (a confirmation of the previous results of Schwartz and Minkov, 2014; Janssen et al., 2015) whereas verbal inflection is not. This finding can be explained from the usage-based perspective on language acquisition: case on nouns in Russian is one of the least transparent morphological categories; it is characterized by high syncretism and multiplicity of manifestations and thus it is more challenging for language acquisition as compared to the iconic and transparent tense-person verb inflection. Children need more input and more time in order to uptake the case forms and acquire their *form-function* meanings. For example, while the inflection *-es'* in *igraes'* "play-2SG.PRES" unambiguously marks the 2nd person present singular, the ending *-i* in *teni* "shadows-NOM/-ACC or shadow-GEN/-DAT/-LOC" can mark plural—nominative or accusative or singular—genitive, dative or locative. These features of the case forms impede their uptake from the input and the establishment of the form-function mapping—thus, children

need more instances of a given category in different contexts in order to be able to identify its meaning and generate a rule. Additionally, the so-called child-directed speech has—in contrast to the language addressed to adult speakers—its specific peculiarities, e.g., a reduced morphological richness, i.e., not all forms of a paradigm are present in child-directed speech, and vocabulary (Hoff-Ginsberg, 1985; Aksu-Koç, 1998; Hoff, 2006 among others). Thus, this positive influence of paradigmatic morphological richness on the speed of case acquisition is weakened by the low degree of morphological richness for nouns (Xanthos et al., 2011) and by syncretism of case inflection. The acquisition process of case inflection on nouns is therefore slowed down and—given the switch to L2 input dominance in bilinguals—cannot progress any more in a sufficient way, i.e. because the L1 input and the frequency of use are not enough for the attainment of the target morphological system.

Another strong pattern in our multiple regression analyses was the notably lower amount of explained variance of morphological skills as compared to lexical skills. This might partially be traced back to the smaller sample size for the morphological data, which causes a moderately lower power in these models. However, another important factor, which was not taken into account, is the interdependence of lexical and grammatical skills. Both domains were highly correlated and based on extensive previous research, allowing a sure conclusion that lexical development is a very important prerequisite for morphological development also in bilinguals (for bilingual children e.g., Simon-Cerejido and Gutiérrez-Clellen, 2009; Kohnert et al., 2010; Blom et al., 2012). Our design did not allow us to include the lexical abilities together with the background factors in a regression model to explain the variance in morphological abilities: due to the wide age range, age and lexical abilities were also highly correlated. Including both together in the regression model would cause multicollinearity. This would therefore be an interesting issue for future studies.

Contrasting the morphological paradigms of case and verb inflections, we found diverging patterns of influencing factors for these morphological categories in L1 acquisition of Russian: case, in our sample, was influenced more by input factors (L2 AoO and L1 use, but also age), and verb inflection more by biological factors (only age and gender). This picture was completed in our longitudinal analyses, which, in general, showed that children showed development in heritage Russian within 1 year in all tested domains. However, this effect was small only for case, in contrast to medium effects for verb inflection as well as the lexical tasks. Although there was an increase, even the oldest children around 5 years of age (AG4 at T2), performed at a very low level in the case test and were as a group far from the full acquisition of this paradigm, showing a fossilization of case and/or, taking another perspective on the heritage grammar—indicating the restructuring of the case system under the conditions of reduced input in bilingual acquisition. This confirms the findings of Gagarina and Reichel (2013), describing case as a vulnerable area in heritage Russian of two Russian-German bilingual children and as an area that more readily undergoes restructuring (the reduction of cases) in the context of input insufficiency and non-sustainability. For verb inflections,

the oldest children in our study performed at a high, close to target, level, which indicates the full or near-full mastery of this paradigm. This difference in the attainment of the target system of verb and noun inflections and in the speed of their acquisition goes back to “a grammatical verb bias (as opposed to a lexical noun bias) in early language development” (Xanthos et al., 2011, p. 472): morphological richness of verbs in child-directed speech is higher as compared to nouns and since it is positively associated with the speed of development in child speech, verb inflections are acquired within a shorter time interval and are more robust to the background factors. The acquisition of case, on the other hand, is more sensitive to input factors—L2 AoO and L1 use—than verb inflection. Generally, bilingual children in our study acquired both morphological domains notably slower than monolingual children, who master the case inflection at age three without errors (Gagarina and Voeikova, 2009, but Janssen et al., 2015 on case processing) and show consistently errorless use of verb tense-person inflection prior to age three (Gvozdev, 1949; Kiebzak-Mandera, 2000; Gagarina, 2003, 2008). “The younger the better” rule of Singleton and Ryan (2004), despite this slower acquisition of verb inflection, appeared to define its robustness against the unfortunate environmental factors impeding attainment of the target L1 morphological system.

## CONCLUSION

This study reported on the differentiated impact of various background factors on L1 acquisition of lexicon and morphology in Russian-German bilinguals in the situation of the change of input dominance from heritage language to L2 upon entering kindergarten. Additionally, it provided new evidence of heritage development of lexicon and morphology—accusative and dative case on nouns and 1st and 2nd person singular present tense inflection on verbs—obtained from the 1-year longitudinal observations of 3- and 4-year-old bilinguals. The background factors were shown to play diverse roles in heritage acquisition of lexicon and morphology. Chronological age was found to impact all investigated domains. Gender impacted all domains except case, and L2 AoO influenced not only case, but also expressive lexicon. The results obtained from the differentiated treatment of input—nuclear family vs. other people—suggested that the evaluation of the nuclear family input is a significant predictor for the acquisition of lexical and case abilities in a heritage language. This does not hold for verbal inflection, which was robust to input effects. Finally, the acquisition of L1 noun and verb morphology showed different patterns of interaction with the background factors—inflection on verbs appeared to be more robust to these factors than case on nouns. The peculiarities of child-directed speech were used to explain this finding:

the morphological richness of the verb paradigm in Russian “stimulated” children to learn it early and fast and with lower input quantity and one form-one meaning relationship made this task easier, so that the shift in the input dominance did not impede the acquisition of verb inflection, which appeared to be rather robust. The acquisition of case in heritage Russian, on the contrary, was found to be less stable and more vulnerable to the background factors. The syncretic nature and non-transparency of case inflections aggravated children’s task of uptaking the noun form, establishing associations between the form and its grammatical function and acquiring the correct contexts of use. In the context of reduced and non-dominant input acquisition of such a category as case becomes a challenging task, which is not fully accomplished and remains unstable till puberty. All in all, this study deepens our knowledge of the development of heritage/home language in Russian-German bilingual children in the context of the shift in input dominance, provides evidence for the differentiated influence of biological and other factors on the acquisition of lexicon and some morphological categories and, finally, enriches our understanding of the multi-faceted process of acquisition of heritage Russian and underlines the decisive role of input in the acquisition of L1.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of German Linguistic Association, Ethics Committee. The protocol was approved by the Ethics Committee, head Prof. Dr. P. Schumacher. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

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

## FUNDING

This study was partially funded by the Berliner Senate (grant to the first author) and German Federal Ministry of Education and Research (BMBF) (grant no. 01UG1411).

## ACKNOWLEDGMENTS

We thank all day-care centers, parents, and children for participation in this study. Special thanks goes to the students and researchers of the Leibniz-ZAS projects, who participated in data collection and analyses. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

## REFERENCES

- Aksu-Koç, A. (1998). The role of input vs. universal predispositions in the emergence of tense-aspect morphology: evidence from Turkish. *First Lang.* 18, 255–280. doi: 10.1177/014272379801805402

- Anstatt, T. (2008). Russisch in Deutschland: Entwicklungsperspektiven. *Bull. Der Deutschen Slavistik* 14, 67–74.
- Armon-Lotem, S., Walters, J., and Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian-Hebrew and Russian-German preschool



- children. *Linguist. Approaches Bilingual*. 1, 291–317. doi: 10.1075/lab.1.3.04arm
- Bayram, F., Rothman, J., Iverson, M., Miller, D., Puig Mayenco, E., Kupisch, T., et al. (2017). Differences in use without deficiencies in competence: passives in the Turkish and German of Turkish heritage speakers in Germany. *Int. J. Bilingual. Edu. Bilingual*. 32, 1–21. doi: 10.1080/13670050.2017.1324403
- 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
- Bornstein, M. H., Hahn, C., and Haynes, O. M. (2004). Specific and general language performance across early childhood: stability and gender considerations. *First Lang.* 24, 267–304. doi: 10.1177/0142723704045681
- Brehmer, B. (2007). “Sprechen sie welch? formen und folgen russisch-deutscher Zweisprachigkeit in Deutschland,” in *Mehrsprachigkeit Bei Kindern und Erwachsenen: Erwerb, Formen, Förderung*, ed T. Anstatt (Tübingen: Narr), 163–186.
- Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingual. Lang. Cogn.* 20, 3–16. doi: 10.1017/S1366728915000863
- Cho, G., Shin, F., and Krashen, K. (2004). What do we know about heritage languages? What do we need to know about them? *Multicul. Edu.* 11, 23–26.
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguist. Approaches Bilingual*. 1, 318–345. doi: 10.1075/lab.1.3.05cho
- Cohen, J. (1988). *Statistical Power Analysis for The Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- De Houwer, A. (2007). Parental language input patterns and children’s bilingual use. *Appl. Psycholinguist.* 28, 411–424. doi: 10.1017/S0142716407070221
- Dieser, E. (2009). *Genuserwerb im Russischen und Deutschen. Korpusgestützte Studie zu ein- und Zweisprachigen Kindern und Erwachsenen*. München: Otto Sagner.
- Döring, N., and Bortz, J. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4th Edn. Berlin: Springer.
- Eisenbeiss, S., Narasimhan, B., and Voeikova, M. D. (2009). “The Acquisition of Case,” in *The Oxford Handbook of Case*, eds A. L. Malchukov and A. Spencer (New York, NY: Oxford), 369–383.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., et al. (2012). Differences between girls and boys in emerging language skills: evidence from 10 language communities. *Br. J. Dev. Psychol.* 30, 326–343. doi: 10.1111/j.2044-835X.2011.02042.x
- Flores, C., Santos, A. L., Jesus, A., and Marques, R. (2017). Age and input effects in the acquisition of mood in Heritage Portuguese. *J. Child Lang.* 44, 795–828. doi: 10.1017/S0305000916000222
- Flores, N., and Garcia, O. (2017). A critical review of bilingual education in the United States: from basements and pride to boutiques and profit. *Annu. Rev. Appl. Linguist.* 37, 14–29. doi: 10.1017/S0267190517000162
- Gagarina, N. (2003). “The early verb development and demarcation of stages in three Russian-speaking children,” in *Development of Verb Inflection in First Language Acquisition: A Cross-Linguistic Perspective*, eds D. Bittner, W. U. Dressler and M. Kilani-Schoch (Berlin: De Gruyter), 131–169.
- Gagarina, N. (2008). *First Language Acquisition of Verb Categories in Russian*. St. Petersburg: Nauka.
- Gagarina, N. (2011). “Acquisition and loss of L1 in a Russian-German bilingual child: A case study,” in *Monolingual and Bilingual Path to Language*, ed S. Cejtin (Moskau: Jazyki slavjanskij kul’tury), 137–163.
- Gagarina, N. (2017). “Monolingualer und bilingualer Erstspracherwerb des Russischen: ein Überblick,” in *Handbuch des Russischen in Deutschland: Migration – Mehrsprachigkeit – Spracherwerb*, eds N. Wulff and K. Witzlack-Makarevich (Berlin: Frank and Timme), 393–410.
- Gagarina, N., Armon-Lotem, S., Altman, C., Burstein-Feldman, Z., Klassert, A., Topaj, N., et al. (2014). “Age, input quantity and their effect on linguistic performance in the home and societal language among Russian-German and Russian-Hebrew preschool children,” in *The Challenges of Diaspora Migration: Interdisciplinary Perspectives on Israel and Germany*, eds R. Silbereisen, P. Titzmann and Y. Shavit (London: Routledge), 63–82.
- Gagarina, N., Armon-Lotem, S., and Gupol, O. (2007). “Developmental variation in the acquisition of L1 Russian verb inflection by monolinguals and bilinguals,” in *On-Line Supplement to the Proceedings of BUCLD 31* eds H. Caunt-Nulton, S. Kulatilake and I. Woo (Somerville, MA: Cascadilla Press), 1–11.
- Gagarina, N., Klassert, A., and Topaj, N. (2010). “Sprachstandstest Russisch für mehrsprachige Kinder/Russian language proficiency test for multilingual children,” in *ZAS Papers in Linguistics 54, Sonderheft* (Berlin: ZAS).
- Gagarina, N., Klassert, A., and Topaj, N. (2015). *Тест по русскому языку 781 для билингвальных детей*. [Russian language proficiency test for multilingual children]. St. Petersburg: Zlatoust.
- Gagarina, N., and Reichel, A. (2013). “Zajcy rasplavjat kogti I Zhish’ ty hir? innovacii rusko-nemeckojazychnyh detej doskol’nogo vozrasta [The hare will melt claws and do you live zdes? The innovations of the preschool Russian-German bilinguals],” in *St. Petersburg School of Ontolinguistics. Festschrift S. N. Ceytlin*, eds T. A. Krugljakova and T. V. Kuz’mina (St. Petersburg: Zlatoust), 195–212.
- Gagarina, N., and Voeikova, M. (2009). “The acquisition of case and number in Russian,” in *Cross-Linguistic Approaches to the Acquisition of Case and Number*, eds U. Stephany and M. Voeikova (Berlin: De Gruyter), 179–215.
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingual. Lang. Cogn.* 12, 213–237. doi: 10.1017/S13667289090004015
- Gutiérrez-Clellen, V. F., and Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Appl. Psycholinguist.* 24, 267–288. doi: 10.1017/S0142716403000158
- Gvozdev, A. (1949). *Formirovanije u Rebenka Grammaticheskogo Stroja Russkogo Jazyka*. [The Construction of the Grammatical System of Russian by the Child]. Moscow: Akademija Pedagogicheskikh Nauk RSFSR.
- 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., Senor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Hoff-Ginsberg, E. (1985). Relations between discourse properties of mothers’ speech and their children’s syntactic growth. *J. Child Lang.* 12, 367–385. doi: 10.1017/S0305000900006486
- Janssen, B., Meir, N., Baker, A., and Armon-Lotem, S. (2015). “On-line comprehension of Russian Case cues in Monolingual Russian and Bilingual Russian-Dutch and Russian-Hebrew Children,” in *BUCLD 39: Proceedings of the 39th Annual Boston University Conference on Language Development*, eds E. Grillo and K. Jepsen (Somerville, MA: Cascadilla Press), 266–278.
- Katičić, A. (2003). “Early verb development in one Croatian-speaking child,” in *Development of Verb Inflection in First Language Acquisition*, eds D. Bittner, W. Dressler and M. Kilani-Schoch (Berlin: de Gruyter), 239–267.
- Kiezbak-Mandera, D. (2000). Formation of the verb system in Russian children. *Psychol. Lang. Commun.* 4, 27–46.
- Klassert, A. (2011). *Lexikalische Fähigkeiten bilingualer Kinder mit Migrationshintergrund. Eine Studie zum Benennen von Nomen und Verben im Russischen und Deutschen*. Dissertation, Philipps-Universität Marburg.
- Klassert, A., and Gagarina, N. (2010). Der Einfluss des elterlichen Inputs auf die Sprachentwicklung bilingualer Kinder: Evidenz aus russischsprachigen Migrantenfamilien in Berlin. *Diskurs Kindheits- und Jugendforschung* 4, 413–425.
- Klassert, A., Kauschke, C., and Gagarina, N. (2014). Object and action naming in Russian- and German-speaking monolingual and bilingual children. *Bilingual. Lang. Cogn.* 17, 73–88. doi: 10.1017/S1366728913000096
- Kohnert, K., and Bates, E. (2002). Balancing bilinguals II: lexical comprehension and cognitive processing in children learning Spanish and English. *J. Speech Lang. Hearing Res.* 45, 347–359. doi: 10.1044/1092-4388(2002/027)
- Kohnert, K., Kan, P. F., and Conboy, B. T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *J. Speech Lang. Hearing Res.* 53, 684–698. doi: 10.1044/1092-4388(2009/08-0126)
- Kupisch, T., and Rothman, J. (2016). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Bilingual.* doi: 10.1177/1367006916654355
- Lein, T., Rothweiler, M., and Hamann, C. (2017). “Factors affecting the performance in child heritage Portuguese in Germany,” in *Proceedings of GALA 2015*, eds J. Choi, H. Demirdache, O. Lungu and L. Voeltzel (Newcastle: Cambridge Scholars Publishing), 147–175.

- Maccoby, E. E. (1966). "Sex differences in intellectual functioning," in *The Development of Sex Differences*, ed. E. E. Maccoby (Stanford, CA: Stanford University Press), 25–55.
- Marchman, V. A., Martínez, L. Z., Hurtado, N., Grüter, T., and Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: comparing direct observation and parent-report measures of dual-language exposure. *Dev. Sci.* 20:e12425. doi: 10.1111/desc.12425
- Maviş, I., Tunçer, M., and Gagarina, N. (2016). Macrostructure components in narrations of Turkish-German bilingual children. *Appl. Psycholinguist.* 37, 69–89. doi: 10.1017/S0142716415000429
- Oller, D. K., and Eilers, R. E. (eds). (2002). *Oller Language and Literacy in Bilingual Children*. Clevedon: Multilingual Matters.
- Oller, D. K., Jarmulowicz, L., Pearson, B. Z., and Cobo-Lewis, A. B. (2011). "Rapid spoken language shift in early second-language learning: The role of peers and effects on the first language," in *Language and Literacy Development in Bilingual Settings*, eds A. Y. Durgunoglu and C. Goldenberg (New York, NY: Guilford Press), 94–120.
- Paradis, J. (2011). Individual differences in child English second language acquisition. *Linguist. Approaches Bilingual.* 1, 213–237. doi: 10.1075/lab.1.3.01par
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Appl. Psycholinguist.* 28, 399–410. doi: 10.1017/S014271640707021X
- Pearson, B. Z., Fernandez, S., Lewedeg, V., and Oller, D. (1997). The relation of input factors to lexical learning by bilingual infants. *Appl. Psycholinguist.* 18, 41–58. doi: 10.1017/S0142716400009863
- Polinsky, M. (2016). Structure vs. use in heritage language. *Linguist. Vanguard* 16, 1–14. doi: 10.1515/lingvan-2015-0036
- Polinsky, M., and Kagan, O. (2007). Heritage languages: in the 'wild' and in the classroom. *Lang. Linguist. Compass* 1, 368–395. doi: 10.1111/j.1749-818X.2007.00022.x
- Rodina, Y., and Westergaard, M. (2017). Grammatical gender in bilingual Norwegian-Russian acquisition: the role of input and transparency. *Bilingual. Lang. Cogn.* 20, 197–214. doi: 10.1017/S1366728915000668
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: romance languages as heritage languages. *Int. J. Bilingual.* 13, 155–163. doi: 10.1177/1367006909339814
- Savickiene, I. (2002). The emergence of case distinctions in Lithuanian. *Linc. Stud. Theoret. Linguist.* 9, 105–115.
- Schwartz, M., and Minkov, M. (2014). Russian case system acquisition among Russian-Hebrew speaking children. *J. Slavic Linguist.* 22, 51–92. doi: 10.1353/jsl.2014.0005
- Simon-Cerejido, G., and Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Appl. Psycholinguist.* 30, 315–337. doi: 10.1017/S0142716409090134
- Singleton, D., and Ryan, L. (2004). *Language Acquisition. The Age Factor*. Clevedon: Multilingual Matters.
- Slobin, D. (2001). "Form-function relations: how do children find out what they are?," in *Language Acquisition and Conceptual Development*, eds M. Bowerman and S. C. Levinson (Cambridge: Cambridge University Press), 406–449.
- Soultanian, N., Mihaylov, V., and Reich, H. H. (2008). "Russisch [Russian]," in *Referenzrahmen zur altersspezifischen Sprachaneignung. Forschungsgrundlagen*, eds K. Ehlich, U. Bredel and H. H. R. Reich [Berlin: Bundesministerium (BMBF) für Bildung und Forschung (Ministry of education and research)], 209–226.
- Thomas, E., Williams, N., Jones, L., Davies, S., and Binks, H. (2014). Acquiring complex structures under minority language conditions: bilingual acquisition of plural morphology in Welsh. *Bilingual. Lang. Cogn.* 17, 478–494. doi: 10.1017/S1366728913000497
- Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: the case of Dutch gender. *Bilingual. Lang. Cogn.* 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. W. P., Van Hout, R. W. N. M., and Schepens, J. J. (2015). The gender gap in second language acquisition: gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PLoS ONE* 10:e0142056. doi: 10.1371/journal.pone.0142056
- Voetikova, M. D., and Gagarina, N. (2002). "Early syntax, first lexicon and the acquisition of case forms by two Russian children," in *Pre- and Protomorphology: Early Phases of Morphological Development in Nouns and Verbs (LINCOM Studies in Theoretical Linguistics 29)*, eds M. D. Voetikova and W. U. Dressler (München: LINCOM Europa), 115–133.
- Wiley, T. (2005). The re-emergence of heritage and community language policy in the U. S. National spotlight. *Modern Lang. J.* 89, 594–601.
- Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., et al. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Lang.* 31, 461–479. doi: 10.1177/0142723711409976
- Yan, S., and Nicoladis, E. (2009). Finding le mot juste: differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingual. Lang. Cogn.* 12, 323–335. doi: 10.1017/S1366728909990101

**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 Gagarina and Klassert. 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.



# Bilingualism in a Case of the Non-fluent/agrammatic Variant of Primary Progressive Aphasia

Nomiki Karpathiou<sup>1,2\*</sup>, John Papatriantafyllou<sup>3,4</sup> and Maria Kambanaros<sup>1</sup>

<sup>1</sup> Department of Rehabilitation Sciences, Cyprus University of Technology, Limassol, Cyprus, <sup>2</sup> Dementia Day Care Center, Athens Alzheimer's Association, Athens, Greece, <sup>3</sup> Memory Disorders Clinic, Athens Medical Center, Athens, Greece, <sup>4</sup> Third Age Center IASIS, Athens, Greece

## OPEN ACCESS

### Edited by:

Judit Gervain,  
Centre National de la Recherche  
Scientifique (CNRS), France

### Reviewed by:

Pauline Pellet Cheneval,  
Université de Genève, Switzerland  
Cornelia Hamann,  
University of Oldenburg, Germany

### \*Correspondence:

Nomiki Karpathiou  
ns.karpathiou@edu.cut.ac.cy

### Specialty section:

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

**Received:** 25 April 2018

**Accepted:** 09 November 2018

**Published:** 26 November 2018

### Citation:

Karpathiou N, Papatriantafyllou J and Kambanaros M (2018) Bilingualism in a Case of the Non-fluent/agrammatic Variant of Primary Progressive Aphasia. *Front. Commun.* 3:52. doi: 10.3389/fcomm.2018.00052

There is a growing body of research on language impairment in bilingual speakers with neurodegenerative diseases. Evidence as to which language is better preserved is rather inconclusive. Various factors seem to influence language performance, most notably age of acquisition, level of proficiency, immersion and degree of exposure to each language. The present study examined fluency, lexical, discourse and grammatical abilities of a Greek-French late bilingual man with the non-fluent/agrammatic variant of primary progressive aphasia (nfvPPA). Speech samples derived from three different narrative tasks in both languages were analyzed using quantitative production analysis (QPA) and fluency measures. The first aim of the study was to compare the participant's connected speech production to that of Greek-speaking normal controls. The second aim was to determine whether Greek (L1) and French (L2) were differentially impaired. To our knowledge, this is the first report of connected speech deficits in a Greek-speaking patient with PPA and the first study which uses QPA to compare L1 and L2 narratives in a bilingual speaker with PPA. Compared to neurologically healthy controls, our participant was impaired in lexical, discourse and grammatical productivity measures, but did not differ in measures of grammatical accuracy. The presence of dysfluencies, reduced speech rate and simplified syntax is consistent with the pattern of impairment reported for the nfvPPA. Results showed that narrative production measures did not differ significantly between languages. However, they suggest a slightly worse performance in his second, non-dominant, language despite a similar pattern of impairment in both languages. Lengthy exposure to L2 and regular activation of L2 through daily use may explain the preservation of discourse abilities in his non-dominant language. This study calls attention to factors such as language dominance, proficiency, patterns of use, and exposure to a language. These factors play a key role in assessing bilingual individuals with PPA and making clinical decisions.

**Keywords:** bilingualism, primary progressive aphasia, PPA, non-fluent, Greek, quantitative production analysis, connected speech, narrative

## INTRODUCTION

The notion of bilingualism refers to the use of two or more languages by an individual in daily life (Grosjean, 1994). First language (L1) and second language (L2) are typically the terms used to characterize languages in respect to their order of acquisition. The terms early and late bilingual classify a person according to the age at which the second language is acquired. Finally, the terms dominant and non-dominant language refer to differences in processing abilities between the two languages and/or in language use. Most researchers agree that both proficiency and use are key contributors to the bilingual experience (Treffers-Daller, 2015).

Bilingualism is a complex construct. Various factors seem to influence language performance in bilingual individuals. Factors related to L2, include age of acquisition, method of acquisition, level of proficiency in the second language and in different modalities (listening, speaking, reading, and writing), similarity to the first language and patterns of language use (e.g., Lorenzen and Murray, 2008; Goral and Conner, 2013; Kambanaros, 2016). In bilingual speakers with an acquired language disorder, language performance in L1 and L2 also depends on the underlying pathophysiology including traumatic brain injury, stroke and neurodegeneration.

Different hypotheses have been put forward to account for language representation in the brain. Evidence comes from electrophysiological investigations and neuroimaging studies of impaired and unimpaired bilingual persons, as well as clinical studies examining the effect of brain damage on language processing in bilingual speakers.

In terms of lexical processing, clinical studies support non-selective lexical access to a multilingual lexicon with shared lexical-semantic representations (e.g., Abutalebi, 2008; Kambanaros, 2016). Parallel lexical-semantic decline in cases of neurodegeneration (Hernández et al., 2008; Costa et al., 2012) or impairment in post-stroke aphasia (Kambanaros and van Steenbrugge, 2006; Kambanaros, 2009, 2010, 2016; Farooqi-Shah and Waked, 2010) are in favor of a common underlying neural network. Neuroimaging studies indicate both shared and separated brain regions for the two languages (Khachatryan et al., 2016).

As for grammar processing, researchers (Paradis, 1994, 2008; Ullman, 2001) have proposed that L1 and L2 are differentially processed as they rely on different cognitive mechanisms: L1 is acquired implicitly through immersion, whereas L2, when it is acquired later in life, explicitly through tuition. Syntactic processes are served by different brain areas, more left anterior (frontal) and subcortical (basal ganglia) regions for L1 and more posterior (temporo-parietal) cortical regions for L2. Others support shared L1 and L2 grammatical representations which are located in common regions (Hartsuiker et al., 2004; Weber and Indefrey, 2009). Evidence from functional neuroimaging studies suggest that L2 processing may become more automatic and converge to the same neural representations of L1 through long exposure to L2 (Abutalebi, 2008). However, differences between first and second language processing have been attributed to cognitive control mechanisms, as the functional demand placed

on these regions is higher for speakers of multiple languages and influenced by factors such as age of acquisition, level of proficiency, and exposure to a language (Abutalebi and Green, 2007; Green and Abutalebi, 2013; Weber et al., 2016).

Evidence from brain imaging studies emphasize the role of L2 proficiency and age of acquisition in interpreting results. In studies where the level of proficiency has been controlled for, there is a higher degree of L1 and L2 overlapping activation for high-proficient than for low-proficient participants (Higby et al., 2013). The dorsolateral prefrontal cortex, anterior cingulate cortex, and right inferior frontal gyrus have been associated with L2 processing in lower proficient bilinguals in a meta-analysis by Sebastian et al. (2011). In another meta-analysis examining the role of age of acquisition in L1 and L2 processing, Liu and Cao (2016) concluded that language networks are more divergent for late bilinguals than for early bilinguals. Regions that were found to be more involved in L2 than in L1 processing were left insula and left middle frontal, inferior frontal and precentral gyri. The left superior frontal gyrus was more recruited by late bilinguals. This result suggests reliance on wider neural resources in the case of late bilinguals.

Primary progressive aphasia (PPA) is a neurodegenerative disease in which language is selectively impaired, at least in the initial stages, providing thus a unique opportunity to study bilingual aphasia and brain representations of language (Filley et al., 2006; Machado et al., 2010). The present study sought to investigate the connected speech deficits in a Greek-French late bilingual person with the non-fluent/agrammatic variant of PPA (nfvPPA). The nfvPPA is characterized by agrammatic production and/or apraxia of speech. Object knowledge and single-word comprehension are usually spared, whereas syntactic comprehension may be impaired. According to the 2011 consensus criteria (Gorno-Tempini et al., 2011), PPA also comprises the semantic (svPPA) and the logopenic (lvPPA) variant. Recently, primary progressive apraxia of speech (PPAOS) has been recognized as a distinct clinical entity (e.g., Duffy et al., 2014). Individuals with PPAOS present with apraxia of speech as their primary deficit and have little or no evidence of aphasia.

Single word production deficits have been extensively examined in PPA and studies of bilingualism. However, connected speech analysis has only recently begun to be systematically studied and has been used only in one study to compare performance in bilingual speakers with PPA (Zanini et al., 2011). The evaluation of connected speech enables a multi-level naturalistic assessment of language production (Marini et al., 2011). All linguistic levels, phonetics, phonology, morphology, syntax, semantics, pragmatics, and discourse can be evaluated when analyzing connected speech samples. Different tasks have been used to elicit speech samples and evidence suggests that they have different specificity for addressing different linguistic levels (Boschi et al., 2017). For example, a picture description task may be more useful in documenting lexico-semantic deficits, whereas story narration tasks favor the evaluation of discourse and syntactic abilities. Spontaneous speech production tasks are more sensitive to morphological, syntactic, and discourse level deficits, as in unconstrained tasks



it is easier for speakers to compensate for their word-finding difficulties.

Deficits in the nfvPPA can arise at the phonetic-phonological level and manifest as a motor speech impairment and/or at the lexical-semantic, morphosyntactic, syntactic, or discourse level and present as agrammatism. Boschi et al. (2017) reviewed the evidence from studies focusing on connected speech deficits in neurodegenerative disorders. People with the non-fluent/agrammatic variant of PPA typically speak at a slower speech rate than healthy controls and make frequent speech sound errors (Ash et al., 2009; Wilson et al., 2010; Rogalski et al., 2011). At the lexical level, an increased number of errors in closed class words has been reported (Knibb et al., 2009; Meteyard and Patterson, 2009; Sajjadi et al., 2012). At the syntactic level, they make grammatical errors (Graham et al., 2004; Sajjadi et al., 2012) and produce simplified sentences with lower number of words per utterance, clauses, verb phrases, and coordinated sentences (Knibb et al., 2009; Wilson et al., 2010; Fraser et al., 2014). Concerning discourse abilities, individuals with the nfvPPA produce a reduced number of words, limited relevant information and they have difficulty maintaining the topic (Graham et al., 2004; Wilson et al., 2010; Sajjadi et al., 2012; Ash et al., 2013; Fraser et al., 2014).

Apart from allowing a multi-level evaluation of the speech and language deficits observed in PPA, connected speech measures enable comparison of patterns of impairment in different languages. For these reasons connected speech analysis has been deemed appropriate for the evaluation of narrative production in our bilingual subject with the nfvPPA. For the structural analysis of connected speech, we used the Quantitative Production Analysis (QPA) (Saffran et al., 1989). QPA was first used to describe agrammatic speech but has been found useful in identifying differences between fluent and non-fluent types of aphasia (e.g., Varkanitsa, 2012) and has been successfully applied in distinguishing normal from aphasic production and differentially diagnosing PPA variants (Wilson et al., 2010). An additional set of fluency measures, error analysis and macrolinguistic measures were also used to allow for a more thorough documentation of the deficits observed in nfvPPA.

A small number of case studies on bilingual speakers with PPA have been published in recent years (Filley et al., 2006; Hernández et al., 2008; Machado et al., 2010; Zanini et al., 2011; Larner, 2012; Druks and Weekes, 2013). Kambanaros and Grohmann (2012) published a case study of a multilingual man with fluent PPA, highly proficient in three languages, Greek, English, and Czech. He was more impaired in L3 than L2 and L1, and more impaired in L2 than in L1. In other words, the extent of impairment in each language was correlated with the order of acquisition. In a short report Machado et al. (2010) presented a Portuguese–French bilingual speaker with PPA. He was impaired in both languages. Performance was overwhelmingly better in his L1 which was also his dominant language. Larner (2012) in another short report, described a Welsh–English speaker who used her L1 in daily communication although L2 was her dominant language. In a more detailed study, Hernández et al. (2008) presented a Spanish–Catalan early bilingual individual

with nfvPPA. They found a naming deficit which was more pronounced for L2 than for L1 at first assessment, but a parallel pattern of decline in both languages, even though L2 deteriorated more rapidly. A grammatical category-specific deficit was present in both languages with an advantage in noun naming over verb naming. A Hungarian–English late bilingual speaker with nfvPPA was reported by Druks and Weekes (2013). Their participant was more impaired in L2 which was his dominant language. A parallel deterioration was found for lexical and grammatical knowledge in L1 and L2. Zanini et al. (2011) described a case of an early Friulian–Italian bilingual woman with nfvPPA. They analyzed her spontaneous speech production and found more phonemic paraphasias, morphological and syntactic errors in L2 than in L1. They reported similar scores for number of dysfluencies, discourse productivity, grammatical productivity, and lexical selection measures (i.e., total words, utterances, subordinate clauses and open-class words) in both languages. Only Filley et al. (2006), who presented a Chinese–English-speaking woman with the logopenic variant of PPA, have reported a non-significant better performance for repetition, naming and conversation tasks, but more phonemic paraphasias, in L2 which was her dominant premorbid language. A parallel pattern of deterioration was observed in both languages. To conclude, most of these studies have found evidence of greater impairment in L2, irrespectively of language dominance and age of acquisition, indicating that L2 may be more vulnerable to degeneration than L1.

In the context of neurodegenerative diseases, there is also a growing body of group studies on language impairment in bilingual speakers with Alzheimer's Disease (AD). The available evidence is mixed. Some studies report parallel deterioration (Salvatierra et al., 2007; Costa et al., 2012; Manchon et al., 2015; Nanchen et al., 2017), while others report differential deterioration of the two languages (Mendez et al., 1999; Gollan et al., 2010). In the study by Gollan et al. (2010), bilingual persons with AD exhibited greater decline in the dominant than the non-dominant language. An opposite pattern was found by Mendez et al. (1999). Based on caregivers' reports, they concluded that the non-dominant language was more affected than the dominant language. Ivanova et al. (2014) found different longitudinal and cross-sectional patterns of decline. The non-dominant language declined more than the dominant language, but differences between patients and controls were greater for the dominant than for the non-dominant language. The authors concluded that both languages are affected by AD with different trajectories of decline over time.

The aim of the present study was 2-fold. First, to provide an account of connected speech deficits in the non-fluent variant of PPA in Greek. The participant's speech and language deficits in his native language were examined by comparing performance on connected speech elicited from a picture description task with speech samples obtained from a healthy control group on the same task. Second, to compare performance in Greek and French and evaluate impairment patterns in both languages connected speech samples from three different narrative tasks in each language were elicited. To our knowledge, this is the first report of connected speech deficits in a Greek-speaking patient

with PPA and the first study which uses QPA to compare L1 and L2 narratives in a bilingual speaker with PPA.

The two languages differ in several respects. Greek is classified as an independent branch within the family of Indo-European languages, whereas French belongs to the Romance branch of the Indo-European family. The components of morphology and syntax are especially relevant to our study. Subject-verb-object (SVO) order is the basic word order in both languages. Word order is flexible in Greek, whereas French has a relatively strict word order. Moreover, Greek is a null subject language, i.e., subjects are not typically expressed when they can be inferred from the context (Roberts and Holmberg, 2010). On the other hand, French is a non-null subject language which requires an explicit subject in a sentence. Regarding morphology, Greek is a highly inflected language, whereas French is considered to be a moderately inflected language. The main difference between the two languages is that in Greek nouns, pronouns, and adjectives are inflected not only for number and gender but also for case. Case in French is expressed using mainly word order and prepositions (Prévost, 2009), although there is a morphological case marking system for weak object pronouns (clitics).

Despite the different linguistic properties of Greek and French, which may result in differences in the narrative measures (e.g., higher proportion of pronouns in French than in Greek because of the mandatory inclusion of subjects in sentences), we predict a similar pattern of impairment in both languages. We also predict that L2, the participant's non-dominant and less proficient language, will be affected to a greater degree compared to L1.

## MATERIALS AND METHODS

### Participant

Participant LJ is a chef in his early sixties, with 6 years of formal education. He is a right-handed late bilingual whose native language (L1) is Greek. At the age of 25, he moved to a French-speaking country and worked as a cook in a French-speaking environment for 7 years. On his return to Greece, he continued to use French (L2) both at work and at home with his wife who is a French native speaker. Details about his language history and proficiency were collected from his wife upon completion of the French version of the Language Experience and Proficiency Questionnaire (Marian et al., 2007) (Table 1). Language dominance was determined based on the reported proficiency and extent of language exposure. Task specific measures of proficiency (for understanding, speaking and reading), across settings measures of language exposure (to family, friend, reading and television) and global measures of these two dimensions were all taken into account in order to ascertain language dominance.

LJ reported a progressive deterioration of speech and language functions. Language impairment was the primary impairment for at least the first two years. LJ was initially assessed 5 years after symptom onset. He received a comprehensive evaluation including case history, neurological examination, and neuropsychological testing coordinated by the second author

**TABLE 1 |** Reported language history and proficiency for participant LJ based on the Language Experience and Proficiency Questionnaire (LEAR-Q, Marian et al., 2007).

Language history measures Languages	L1 history Greek	L2 history French	L3 history English	Range
Order of proficiency	1	2	3	
Order of acquisition	1	3	2	
Identification with culture <sup>a</sup>	10	6	1	0–10
Current exposure	46%	46%	8%	
Preference for reading	80%	20%		
Preference for conversing	40%	40%	20%	
<b>REPORTED PROFICIENCY<sup>b</sup></b>				
Understanding	4	5		0–10
Speaking	5	5		0–10
Reading	2	0		0–10
<b>AGE MILESTONES (YEARS)</b>				
Started learning		25		
Attained fluency		29		
Started reading	6	25		
Became fluent reading		n/a		
<b>IMMERSION DURATION (YEARS)</b>				
Country	53	7		
Family	53	36		
School/Job	53	36		
<b>CONTRIBUTION TO LANGUAGE LEARNING<sup>c</sup></b>				
From family	10	10		0–10
From friends	0	8		0–10
From reading	0	0		0–10
From TV	2	5		0–10
From radio	0	0		0–10
From self -instruction	0	1		0–10
<b>EXTENT OF LANGUAGE EXPOSURE<sup>d</sup></b>				
To family	10	10		0–10
To friends	10	7		0–10
To reading	1	1		0–10
To TV	7	3		0–10
To radio	0	0		0–10
Self -instruction	0	0		0–10
<b>SELF -REPORTED FOREIGN ACCENT<sup>e</sup></b>				
Perceived by informant	2	5		0–10
Identified by others	5	5		0–10

<sup>a</sup>Range, 0 (none) to 10 (complete); <sup>b</sup>Range, 0 (none) to 10 (perfect); <sup>c</sup>Range, 0 (not a contributor) to 10 (most important contributor); <sup>d</sup>Range, 1 (never) to 10 (always); <sup>e</sup>Range, 0 (none) to 10 (pervasive).

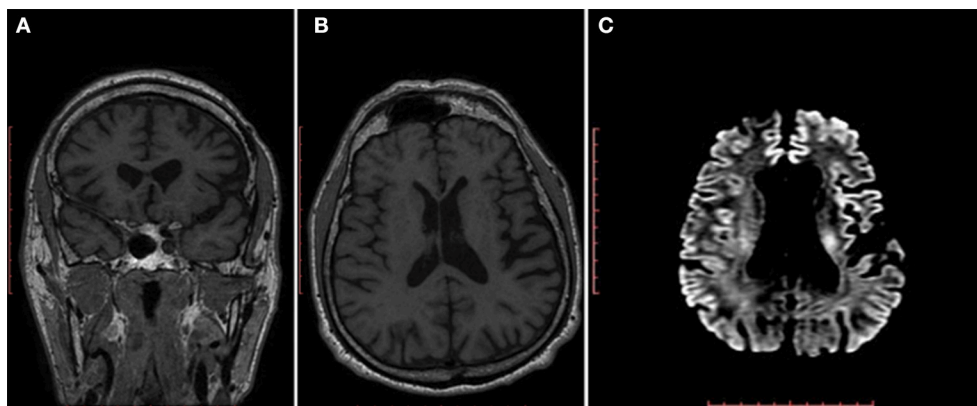
who is a psychiatrist specialized in memory disorders with extensive experience working with patients with degenerative diseases. He was referred for speech and language evaluation and completed an initial language assessment performed by the first author in Greek. He was diagnosed with PPA, as neuroimaging results ruled out other causes of focal brain damage and extensive white matter disease (see Figure 1) and was given a clinical diagnosis of non-fluent/agrammatic PPA according to current criteria (Gorno-Tempini et al., 2011). There

were no signs of limb apraxia, tremor, dystonia and myoclonus. There was a very mild hypertonicity on the right side, as well as reports of becoming more suspicious of others. His speech was slow with word finding problems, hesitations, pauses, and sound errors. Motor speech evaluation determined the presence of apraxia of speech with slow overall rate, deliberate, slowly sequenced speech sequential motion rates in comparison to speech alternate motion rates, imprecise articulation with sound distortions, a tendency to equalize stress across syllables, false starts and restarts and sound and syllable repetitions. Dysarthria, most probably spastic, was present, but less severe than apraxia of speech. LJ had spared knowledge of objects and word recognition. A mild difficulty comprehending syntactically complex sentences was revealed in formal testing. His consensus score on the Progressive Aphasia Severity Scale (PASS) (Sapolsky et al., 2010) was 7 (see **Table 2**). Background linguistic and neuropsychological evaluation results are presented in **Table 3**.

Prior to testing for the present study, LJ had received speech and language therapy for approximately 4 months. Intervention included partner education, script training (Youmans et al., 2005) of telephone conversations with clients and techniques based on

the “Oral Reading for Language in Aphasia” treatment program (Cherney, 2010) that addressed production of multisyllabic words, as well as reading and auditory comprehension. Treatment was delivered in Greek.

The present study was conducted 9 months after the initial evaluation (5 years and 9 months after the reported onset of the disease) and 3 months after the last therapy session. At the time of the study, LJ had a FTLD-modified CDR sum of boxes score of 9 (MMSE = 17/30). The Montreal Cognitive Assessment (MOCA) was administered both in Greek and French. He received a score of 18/30 in Greek and 20/30 in French (one additional point in visuospatial/executive function and one in memory). He generated 2 words in the phonemic verbal fluency task and 5 words in the semantic task (animals) and obtained a score of 3 on the forward digit span and 0 on the backward digit span. There was also a parallel deterioration of motor skills. These results suggest a deterioration in cognitive function, especially in the domain of executive function and progression of the nvPPA to a corticobasal syndrome. Corticobasal syndrome can overlap clinically and pathologically with PPA and many cases initially classified as nvPPA, meet the criteria for corticobasal syndrome at a later



**FIGURE 1** | Coronal T1-weighted (A), axial T1-weighted (B), and axial diffusion-weighted (C) brain imaging at initial assessment showing left perisylvian atrophy.

**TABLE 2** | Consensus score on the Progressive Aphasia Severity Scale (PASS) at initial evaluation.

PASS Domains	Normal	Quest/ble very mild	Mild	Moderate	Severe
Articulation	0	0.5	1	2	3
Fluency			1		
Syntax and grammar			1		
Word retrieval - expression			1		
Repetition	0				
Auditory comprehension		0.5			
Single word comprehension	0				
Reading		0.5			
Writing			1		
Functional communication			1		

Severity (Sum of boxes), 7.

**TABLE 3 |** Background neuropsychological assessment results.

Area of testing and tests	Score (correct)
<b>GENERAL COGNITIVE MEASURES</b>	
MMSE	28/30
ACE-R	86/100
Attention	18/18
Memory	26/26
Fluency	5/14 *
Language	25/26
Visuospatial abilities	12/16 *
<b>EXECUTIVE FUNCTIONING</b>	
Frontal Assessment Battery (FAB)	12/18
<b>VISUOSPATIAL PERCEPTION</b>	
Benson figure test—Copy condition	15/17
<b>VISUAL MEMORY</b>	
Benson figure test—Delayed recall condition	17/17
<b>MOOD</b>	
GDS-SF	3/15
<b>IDEOMOTOR APRAXIA</b>	
WAB	58/60
<b>REPETITION</b>	
Informal (based on WAB)	95/100
<b>NAMING</b>	
Boston Naming Test, BNT-SF	11/15 *
<b>LANGUAGE COMPREHENSION</b>	
Vocabulary (PPVT-32)	19/32
Auditory comprehension-words (BDAE-SF)	16/16
Sequential commands (BDAE-SF)	10/10
Written sentences/passages (BDAE-SF)	4/4
Written story (BDAE-SF)	3/3
Grammaticality judgment—morphology (Fyndanis et al., 2013)	77/80
Syntactic comprehension (BDAE-3)	8/10
<b>OBJECT SEMANTICS</b>	
Pictures (PPTT-SF)	14/14
<b>READING EFFICIENCY (Simos et al., 2013)</b>	
Real words	16 in 45s *
Pseudowords	13 in 45s
<b>WRITING</b>	
Words (Informal)	7/20
Non-words (Informal)	14/14
Words (BDAE-SF)	8/9
Written picture description (BDAE-SF)	4/11 *
<b>MOTOR SPEECH EVALUATION (Wertz et al., 1984)</b>	
Apraxia of speech rating	3/7
Dysarthria rating	1/7

\*Significant impairment (>2 standard deviations below the normative mean); MMSE, Mini Mental State Examination (Fountoulakis et al., 2000); ACE-R, Addenbrooke's Cognitive Examination-Revised (Konstantinopoulou et al., 2011); GDS-SF, Geriatric Depression Scale-Short Form (Fountoulakis et al., 1999); WAB, Western Aphasia Battery; BDAE-SF, Boston Diagnostic Aphasia Examination Short form (Goodglass et al., 2013); PPVT, Peabody Picture Vocabulary Test (Simos et al., 2011); PPTT-SF, Pyramid and Palm Trees Test-Short Form (Breining et al., 2015).

stage (Grossman, 2010; Duffy et al., 2014; Leyton and Ballard, 2016; Santos-Santos et al., 2016).

The study was approved by the ethics committee of the Athens Alzheimer's Association. The research was conducted in

accordance with the latest version of the Declaration of Helsinki. LJ was informed about the purpose and procedures of the study and gave written consent for participating in the study, as well as for the recording and publication of his clinical data. Both LJ and his wife gave written informed consent for the publication of this manuscript. The initials LJ are fictional.

## Elicitation and Transcription of Speech Samples in L1 (Greek) and L2 (French)

Three different speech samples were collected in both Greek and French, under 3 conditions: a picture description task ("cookie theft," from Boston Diagnostic Aphasia Examination, BDAE), a story retell task (the dog story protocol from the Multilingual Assessment Instrument for Narratives, MAIN, Gagarina et al., 2012, 2015) and a semi-spontaneous speech task where LJ was asked to talk about his job. Interruptions and questions by the examiner (first author) were kept to a minimum. The examiner is a monolingual Greek-speaking clinician who is also a proficient speaker of French. Samples were collected in 4 sessions, first for the Greek language and 2 weeks later for French. All samples were audio-recorded.

Speech samples were transcribed orthographically using ELAN (Sloetjes and Wittenburg, 2008). Phonological paraphasias unintelligible or incomprehensible words were transcribed phonetically using the International Phonetic Alphabet. Dysfluent variables, such as silent and filled pauses, sound errors, repetitions, and false starts were also coded.

## Quantitative Analysis of Speech Samples

Speech samples were analyzed following the procedures described by Saffran et al. (1989) for quantitative production analysis (QPA) (Saffran et al., 1989; Berndt et al., 2000; Rochon et al., 2000). The QPA procedures were followed for all samples, with the exception of the direct discourse utterances produced in the story retell task, which contrary to the QPA instructions were not excluded, as these structures were modeled in story-telling. Narrative samples were formed by extracting comments on the narrative, direct responses to the examiner, repetitions of the examiner's utterances, stylistic and dysfluent repetitions, subsequently repaired utterances and discourse markers. The narrative samples were then segmented into utterances based on semantic, syntactic, and prosodic information. Utterances and narrative words were used in subsequent analysis.

The QPA summary measures were classified into four categories: discourse productivity, sentence productivity, grammatical accuracy, and lexical selection (Gordon, 2006). A set of additional measures were used to quantify dysfluent speech and narrative variables.

## Speech Rate and Other Fluency Variables

Speech rate for each sample was calculated by dividing total completed words by sample duration in minutes. Samples were timed, and total time duration was computed by subtracting the examiner's interjections.

Pauses longer than 1 s were coded according to QPA instructions and counted for the calculation of the pause frequency measure. However, a threshold of 0.250 ms was used



in the calculation of pause duration (De Jong and Bosker, 2013) and speaking time was calculated by subtracting silent pausing time from total time in order to control for the effect of pauses. Articulation rate was computed by dividing total completed words by speaking time.

Speech sound errors included distortions, which were defined as phonetic errors resulting in distorted phonemes, and phonological paraphasias defined as words with non-distorted phonemic insertions, deletions, or substitutions. Whole-word immediate repetitions were counted as dysfluent repetitions. Words or phrases repeated later in the narratives were counted as speech repairs. Partially produced words were coded as false starts and small words, such as “eh,” as filled pauses.

Speech samples were of different duration and direct comparison of the aforementioned frequency measures was not possible. Thus, these measures were calculated as proportions of total words produced. They were also corrected for speaking length by dividing dysfluency counts by speaking time (De Jong, 2016).

### Discourse Measures

QPA discourse productivity measures included speech rate, number of narrative words, and proportion of narrative to total words produced, as a measure of discourse efficiency.

An additional discourse variable, Guiraud's index (the square root variant of Type-Token Ratio, TTR) was also measured. Guiraud's index is a measure of lexical richness that is less affected by sample size/length in comparison to TTR (Van Hout and Vermeer, 2007). This was derived by dividing the number of unique words (types) by the square root of narrative words (tokens). Number of unique words (types), lemmas, and utterances are also reported.

### Lexical Measures

Grammatical category class (closed/open class, nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions) was coded for each narrative word. Their proportion was calculated by dividing the number of words in each category by the number of narrative words. Nouns, verbs and adjectives were considered as open class. All other words were counted as closed class. Proportion of verbs to nouns and verbs was also computed. Proportion of pronouns was derived by dividing the number of pronouns by the total number of nouns and pronouns.

Finally, mean log word frequency of open class words was calculated for each narrative sample. Calculations were based on data about word frequencies per million taken from the “ILSP PsychoLinguistic Resource” for the Greek language (Protopapas et al., 2012) and “Lexique” for the French language (New et al., 2001).

### Grammatical Measures

QPA sentence productivity measures encompass proportion of words in sentences, mean utterance length (in words), median utterance length (in words), sentence elaboration index (number of open class words per phrase for noun and verb

phrases), and an embedding index (proportion of embeddings to sentences).

QPA grammatical accuracy measures consist of proportion of well-formed sentences, verb inflection index (proportion of inflectable verbs inflected) and determiner index (proportion of determiners produced in obligatory contexts). The auxiliary complexity index, a measure of morphological complexity of the main verb indicating change from its base form, was also calculated.

## Macrolinguistic Analysis (MAIN)

Narrative assessment focused on the analysis of microlinguistic aspects of language production. Macrolinguistic aspects were addressed for the “Dog story” retell task with the story structure score and the structural complexity measures proposed by MAIN (Gagarina et al., 2012, 2015). Although the MAIN was originally designed to assess narrative skills of bilingual children, it is controlled for macro-and microlinguistic features across Greek and French. As there is no other standardized procedure for adults, it was deemed appropriate for comparing story retell abilities in both languages.

The “Dog story” starts with a setting statement and consists of three short episodes. Each episode consists of an initiation, a goal, an attempt, an outcome and a reaction statement. Credit is given for the production of each initiation, goal, outcome, reaction when computing the story structure score.

Five measures of structural complexity, included in the MAIN, were calculated: number of sequences where an attempt and outcome statement has been generated (but no goal), number of single goal statements, number of incomplete episodes which they include a goal and an attempt statement sequences, number of incomplete episodes which they include a goal and an outcome statement, and number of complete episodes which include all three goal-attempt-outcome components. Comprehension of the story structure was also assessed by means of questions targeting the main macrostructure components.

## Error Analysis

The following type of errors were also identified and measured as a proportion of narrative words.

- Syntactic errors were recorded when LJ produced ungrammatical sentences.
- Morphological errors, affecting articles, nouns, adjectives, and verbs, were counted separately.
- Semantic errors included selections that were semantically inappropriate for the context.
- Code switching errors were defined as words produced in languages other than the target language (number of tokens not in the target language).

Some morphological errors in L2 (article-noun gender agreement) occurred with the same nouns. These persistent errors were not included in individual error counts but contributed to the calculation of the total number of errors.

## Inter-rater Reliability

Analysis of 30% of the Greek speech samples was completed by 2 additional raters both native speakers of Greek with some linguistic training. Spoken word interrater reliability ranged from 90 to 95%. A consensus for each point of disagreement was reached through a discussion between the raters.

## Control Group for QPA

QPA measures for the picture description task in Greek were compared to the measures of a control group included in a previous study by Varkanitsa (2012). Varkanitsa used the QPA protocol in order to compare the connected speech of Greek-speaking persons with aphasia following stroke to that of neurologically healthy adults. The same picture description task was used in the present study to elicit speech samples. Taking into account the fact that in Greek isolated verbs may constitute grammatical utterances, Varkanitsa categorized utterances as “utterances with verb,” “utterances without verb” and “single-word utterances.” The QPA protocol was applied without other modifications. The control group consisted of six normal native Greek speakers (3 males and 3 females) with a mean age of 61.17 (SD = 5) years and a mean of 9 (SD = 4.15) years of education.

There was no control group for QPA measures in French, as we did not have access to a French-speaking population and published studies, which have applied QPA in French-speaking individuals, have not used the same methodology. For this reason, our analysis focused on the pattern of deficits observed in the two languages. Moreover, careful consideration was given to cross-linguistic differences.

## Statistical Analysis

LJ's narrative scores for the picture description task in Greek were compared to the scores of a neurologically healthy control group (Varkanitsa, 2012). *T*-values were calculated using Crawford and Howell's method which enables the comparison of performance of a single subject with that of a small control sample (Crawford and Garthwaite, 2012). Differences between LJ's performance in Greek (L1) and French (L2) were calculated using the Wilcoxon signed-rank non-parametric test for related samples because of the small sample size. Finally, scores from both languages were collapsed and correlations between errors and fluency, lexical productivity, grammatical accuracy, and productivity measures were calculated using the non-parametric Kendall's tau-b correlation coefficient due to the limited number of samples used in the analysis.

## RESULTS

### QPA Measures for the Picture Description Task in Greek—Comparison to Healthy Subjects

LJ's scores for the picture description narrative in Greek are presented in Table 4. His speech rate was slow, 40.37 words per minute. In the picture description task, he made two syntactic errors. Both errors involved the omission of obligatory post-verbal arguments. He also made speech errors. Dysfluencies included silent pauses, filled pauses, false starts,

**TABLE 4 |** LJ's scores, control group median and standard deviation values and Crawford-*t* values.

Spoken language measures	LJ	Controls <sup>a</sup> ( <i>n</i> = 6)	<i>t</i> -values <sup>b</sup>
		Median (SD)	
Proportion of closed class words	0.52	0.53 (0.04)	−0.23
Proportion of nouns	0.17*	0.25 (0.03)	−2.47
Proportion of adjectives	0.04	0.02 (0.01)	1.85
Proportion of prepositions	0.02	0.06 (0.02)	−1.85
Proportion of adverbs	0*	0.07 (0.02)	−3.24
Proportion of pronouns	0.14**	0.06 (0.01)	7.41
Proportion of verbs	0.31*	0.20 (0.04)	2.55
MLU	4	5.41 (1.08)	−1.21
Elaboration index	1	2.43 (1.43)	−0.93
Embedding index	0.36	0.36 (0.17)	0.00
Number of narrative words	48*	127 (35)	−2.09
Proportion of sentences	0.92	0.79 (0.11)	1.09
Proportion of utterances without verbs	0.08	0.19 (0.11)	−0.93
Proportion of single-word utterances	0.17**	0.00 (0.02)	7.87
Proportion of well-formed utterances	0.75	0.96 (0.65)	−0.30
Auxiliary complexity index	0.64	0.30 (0.27)	1.17

<sup>a</sup>Control group values are taken from Varkanitsa (2012); <sup>b</sup>One-tailed (\**p* < 0.05; \*\**p* < 0.01).

sound distortions, and repetitions (23, 20, 3, 2, and 1%, respectively, of total words produced). Compared to the control group, LJ used less narrative words [ $t_{(5)} = -2.089$ ,  $p < 0.05$ ] and more single word utterances [ $t_{(5)} = 7.869$ ,  $p < 0.0005$ ] to describe the picture. Sentence productivity measures (mean length of utterance, elaboration index and embedding index) did not differ from controls. LJ produced less nouns [ $t_{(5)} = -2.468$ ,  $p < 0.05$ ] and adverbs [ $t_{(5)} = -3.240$ ,  $p < 0.025$ ]. On the other hand, he produced more pronouns [ $t_{(5)} = 7.406$ ,  $p < 0.0005$ ] and verbs [ $t_{(5)} = 2.546$ ,  $p < 0.05$ ] than the control speakers.

## Comparison of L1 and L2

Statistical analysis using the Wilcoxon signed-rank test revealed that the connected speech measures used to quantify speech production in L1 and L2 did not differ significantly across languages.

## Fluency Measures

The mean duration of narratives was 2.24 (SD = 0.09) minutes for L1 and 3.76 (SD = 1.86) for L2. Pause duration, for pauses >0.250 ms, was 0.74 (SD = 0.11) minutes for L1 and 1.27 (SD = 0.36) for L2. Speaking time was 1.5 (SD = 0.08) minutes for L1 and 2.49 (SD = 1.58) for L2. Speech rate was faster for L2 than for L1, 44.10 (SD = 5.96) and 38.24 (SD = 2.52) words per minute (wpm), respectively. Similar results were noted for articulation rate: 73.00 (SD = 19.09) wpm for L2 and 57.43 (SD = 6.93) wpm for L1. However, these differences were not statistically significant.

Dysfluencies included silent pauses, fillers, false starts, distortions and immediate repetitions of whole words and in

particular closed class words. The different types of dysfluencies are presented in **Figure 2**.

Although differences between languages did not reach statistical significance, there is a trend toward making more repetitions in L2, 0.040 (SD = 0.012) than in L1, 0.004 (SD = 0.007).

### Discourse Measures

LJ produced longer narratives in L2 than in L1, 94.67 (SD = 68.06) words and 16.33 (SD = 10.12) utterances vs. 53.00 (SD = 8.66) words and 11.00 (SD = 3.61) utterances, respectively. Differences were not significant. From the narrative words, 47.00 (SD = 11.36) words in French and 34.33 (SD = 3.22) words in Greek were unique. Proportion of narrative to total words produced was 0.61 (SD = 0.07) in L1 and 0.55 (SD = 0.21) in L2.

### Lexical Measures

Regarding word class production, significant differences between L1 and L2 were not found. However, LJ produced more closed class words and pronouns in L2 compared to L1 narratives. In French, the proportion of closed class words was 0.56 (SD = 0.01), while in Greek, it was 0.49 (SD = 0.03). The proportion of pronouns was 0.22 (SD = 0.04) in L2, as opposed to 0.12 (SD = 0.04) in L1. LJ produced personal, demonstrative, indefinite and interrogative pronouns. In L1, all demonstrative pronouns (37.5%) were used as subjects, whereas all the rest, including personal pronouns (50%) in their weak form, were produced as object pronouns (62.5%). Of all the pronouns produced in L2, 87.7% were personal pronouns and 8.78% demonstrative. Ninety-Four percent of the personal pronouns were used in their strong form and the remaining 6% in their

weak form. In L2, 87.7% of the pronouns produced were subject pronouns and 12.3% object pronouns.

LJ used more nouns per narrative words in Greek, ranging from 0.17 to 0.31 with a mean of 0.26 (SD = 0.08), in comparison to his L2 in which the proportion of nouns was 0.17 (SD = 0.04), ranging from 0.12 to 0.20. The proportion of verbs produced did not differ across languages (see **Table 5**).

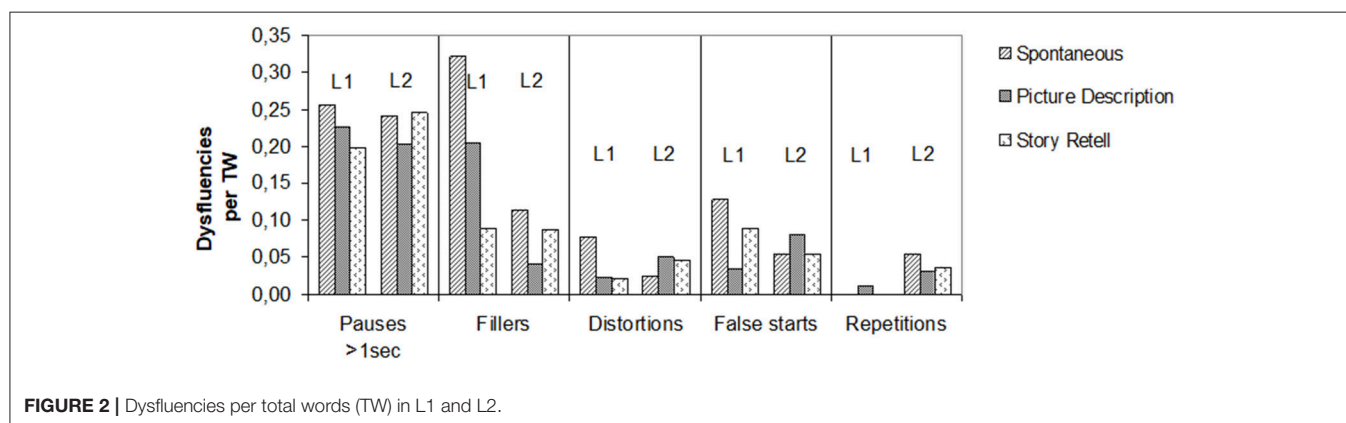
LJ used more high frequency words in French than in Greek narratives. The mean logarithmic frequency of French open class words was 1.71 (SD = 0.17), as opposed to 1.392 (SD = 0.15) for Greek words. This difference was not statistically significant.

### Grammatical Productivity and Accuracy Measures

With regard to measures associated with grammatical production, no statistically significant differences were found between L1 and L2. Mean and median length of utterance in words was 5.12 (SD = 1.53) and 4.17 (SD = 1.04) for Greek and 5.58 (SD = 0.59) and 5.00 (SD = 1.00) for French respectively. LJ performed more poorly in L2 than in L1 as far as the proportion of embedded clauses is concerned (0.19 (SD = 0.08) for L2 and 0.34 (SD = 0.17) for L1).

### Macrolinguistic Measures for MAIN

The MAIN story structure and comprehension scores were 7/17 and 10/10 in L1 and 9/17 and 7/10 in L2, respectively. LJ produced one single goal statement in both languages. In French, he also used a sequence with an attempt and outcome statement. Neither incomplete episodes with a goal and an attempt/outcome statement nor complete episodes (with all three components) were present in his narratives.



**TABLE 5** | Proportion of nouns, verbs and pronouns per narrative words (NW) produced in personal narrative (task1), picture description (task 2) and story retell (task 3) in L1 and L2.

Proportion per NW	L1				L2			
	Task 1	Task2	Task 3	Total Mean (SD)	Task1	Task 2	Task 3	Total Mean (SD)
Nouns	0.31	0.17	0.29	0.26 (0.08)	0.18	0.12	0.20	0.17 (0.04)
Verbs	0.21	0.31	0.24	0.25 (0.05)	0.20	0.30	0.18	0.23 (0.07)
Pronouns	0.08	0.15	0.14	0.12 (0.04)	0.20	0.26	0.19	0.22 (0.04)

## Error Analysis

Systematic errors involving article gender agreement in L2 were excluded from analysis.

LJ made more morphological and semantic errors per narrative words in L2, 0.031 (SD = 0.025) and 0.022 (SD = 0.006), respectively, than in L1, 0.014 (SD = 0.024) and 0.005 (SD = 0.009), respectively. These differences were not statistically significant. Syntactic errors were stable across languages, 0.026 (SD = 0.016) for L1 and 0.023 (SD = 0.008) for L2.

Code switching was evident in one speech sample (spontaneous narrative) in French. LJ produced 11 out of the 166 complete words in Greek.

## Correlational Analysis

We undertook correlational analyses between errors and connected speech measures. Syntactic errors were significantly correlated with the total number of dysfluencies per total words ( $\tau b = 0.733$ ,  $p = 0.039$ ), whereas morphological errors with the distortions produced per articulation minute ( $\tau b = 0.966$ ,  $p = 0.007$ ). Finally, there was a positive correlation between semantic errors and number of complete words ( $\tau b = 0.867$ ,  $p = 0.015$ ).

## DISCUSSION

The present study examined fluency, lexical content, discourse and the grammatical abilities of a Greek-French late bilingual man with non-fluent/agrammatic PPA by analyzing speech samples derived from three different discourse tasks in both languages.

The first aim of the study was to compare the participant's performance to normal controls in L1. Compared to Greek-speaking neurologically healthy individuals, LJ was impaired in discourse and grammatical productivity measures, but did not differ in measures of grammatical accuracy. At the lexical level, there were some significant differences in the proportion of grammatical class words produced. In particular, LJ produced more verbs and pronouns, but less nouns and adverbs. However, proportion of closed class words was normal.

The second aim of the study was to determine whether or not L1 and L2 were differentially impaired. Results showed that discourse production measures did not differ significantly between languages. These findings indicate that both languages were similarly affected.

## Comparison With Healthy Controls in L1 (Greek)

LJ produced a smaller number of narrative words, shorter utterances and simplified sentences compared to controls, as indicated by the MLU, proportion of single-word utterances and elaboration index measures. Production of embedded clauses was at the same level with the control group. The auxiliary complexity index, a measure of verb morphological complexity, was slightly higher for LJ than controls. However, the proportion of single-word utterances is the only grammatical productivity measure that reached statistical significance. Grammatical accuracy did not differ between LJ and neurologically healthy individuals,

even though he produced a lower proportion of well-formed utterances. In the picture description task in Greek, LJ made two errors. Both errors were syntactic in nature and involved the omission of obligatory post-verbal arguments. Taken together, these results indicate an impairment at the discourse and grammatical productivity levels.

Fluency, as measured by speech rate and frequency of dysfluent errors, is another area that was affected. Although we had no control data for the fluency variables, slow speech rate and high proportion of pauses and fillers corroborate reduced fluency. Indicatively, a normal speech rate of 143.70 (SD = 23.40) wpm has been reported for the "cookie theft" description task in a study by Fyndanis et al. (2013). The measure was based on three neurotypical Greek-speakers with a mean age of 58 (SD = 9.64) years. The presence of distortions and false starts indicate an underlying motor speech problem, apraxia of speech in particular (Ogar et al., 2007; Wilson et al., 2010).

Differential impairment of nouns and verbs has been reported in aphasia resulting from stroke and PPA. In particular, disproportionate impairment of naming actions is commonly associated with non-fluent types of aphasia (Kambanaros, 2010) and greater verb naming impairment has been found in nfvPPA (Hillis et al., 2006; Ash et al., 2009; Thompson et al., 2012). Even though LJ used more verbs than nouns during the picture description task in Greek, indicating an opposite pattern of noun-verb dissociation, mean noun-verb ratio from all three Greek speech samples was within normal limits. In fact, higher proportion of verbs seems to be task-related, as disproportionate production of verbs was evident in both languages for the picture description task only. Normal ratios of nouns to verbs in connected speech of individuals with the nfvPPA have been reported in several studies (Graham et al., 2004; Knibb et al., 2009; Meteyard and Patterson, 2009; Fraser et al., 2013; Marcotte et al., 2017).

LJ also used more pronouns in Greek (0.14 per narrative words, 80% demonstrative, 20% personal) than the control group in the picture description task. Increased proportion of pronouns has been found in svPPA and it has been suggested that it may indicate lexical retrieval deficits, vague, or non-specific speech (Kavé et al., 2007; Meteyard and Patterson, 2009; Wilson et al., 2010; Fraser et al., 2014). Nevertheless, all the pronouns used by LJ had clear referents. Furthermore, all the demonstrative pronouns were used in the subject position of sentences. In a null subject language like Greek, demonstrative pronouns may be used as subjects to place additional emphasis on the referent. The production of overt subjects in Greek could reflect the influence of the syntactic properties of the participant's L2 on his L1. Syntactic attrition effects have been reported in the production of preverbal subjects in a group of Greek (L1) speakers, highly proficient in English (L2) (Tsimpli et al., 2004). However, in the personal monolog LJ produced a substantially lower proportion of pronouns (0.08 per narrative words) than in the two picture-based tasks. This most probably suggests that LJ was using demonstrative pronouns to direct the attentional focus to the referent in the depicted scenes. It must be noted that, although the examiner's instruction for the picture description task was "tell me everything you see going on in this picture," for the story



retell task, the instructions focused on the story itself, not the pictures (“Can you tell me the story?” “Tell me more.”). Picture-based tasks have been reported to result in the production of descriptions of the depicted items, rather than narrative samples (Bryant et al., 2016).

Wilson et al. (2010) used a similar methodology to ours by combining QPA and fluency measures to analyze narrative production of 50 English-speaking individuals with PPA. Speech samples were elicited through a picture description task. They found that their nfvPPA group compared to normal controls spoke slower, produced less words and their samples were of longer duration. All nfvPPA participants made distortions and more filled pauses than controls. Their mean length of utterances and number of embeddings were significantly reduced. In respect to the other variants of PPA, the authors concluded that the presence of distortions was the most informative measure for distinguishing between the nfvPPA and lvPPA. Additional measures that may assist in differentially diagnosing these subtypes are proportion of verbs and number of embeddings, which are higher in the lvPPA. Faster speech rate, less distortions, higher proportion of pronouns and verbs and nouns of higher frequency were found in the svPPA compared to the nfvPPA.

LJ's scores support the pattern of impairment reported for the nfvPPA variety. In comparison to neurotypical controls, he made distortions, spoke slower, produced less words and more single word utterances. Although agrammatism has been described as a core characteristic of this variant (Ash et al., 2009; Thompson et al., 2012), grammatical deficits may not be the primary feature of nfvPPA (Graham et al., 2004; Patterson et al., 2006; Wilson et al., 2010). In a recent study, Graham et al. (2016) evaluated fluency and grammatical production in nine individuals with nfvPPA. They reported that frank agrammatism was not always present and reviewing the literature they pointed out that grammatical abilities in persons with the nfvPPA show a high degree of variability. Nevertheless, researchers have consistently reported reduced speech rate, as well as simplified syntax and shorter utterances in connected speech in comparison to healthy controls (Ash et al., 2009, 2010, 2013; Knibb et al., 2009; Wilson et al., 2010; Marcotte et al., 2017).

## Comparison of L1 and L2

The observed differences between L1 and L2 did not reach statistical significance, contrary to our hypothesis. This may be due to the small sample size of linguistic data or the between-task variability. Alternatively, findings may be interpreted as indicating a similar degree of impairment in both languages. Before commenting on this finding, there are some trends in the results that are worth mentioning.

The total number of dysfluencies was similar across languages. However, LJ produced more immediate repetitions in L2 than in L1. He repeated mostly personal pronouns at the beginning of utterances, or after silent pauses. In French, personal pronouns are short monosyllabic words, like “je” /ʒə/ (I), “il” /il/ (he), “elle” /el/ (she). In this case, repetitions seem to be a manifestation of speech initiation difficulty and may be considered as false starts. They were counted separately, though, because of the definition we used; only partially repeated words were counted as false

starts. Had they been clustered together, we would not have found a differential pattern of impairment in L1 and L2 for repetitions nor false starts.

LJ produced more filled pauses in L1 than in L2. Pauses are considered to be indicative of cognitive or linguistic processing difficulties (Krivokapi, 2007; Davis and MacLagan, 2009). In PPA, pauses have been associated with discourse, syntactic and motor speech planning, as well as word retrieval difficulties (Wilson et al., 2010; Mack et al., 2015). Given the fact that the underlying conceptualization process is the same in both languages, this finding cannot be attributed to different level of discourse processing abilities in L1 and L2. Results from the MAIN support a similar pattern of structural discourse deficits in both languages. Similarly, it cannot be attributed to differences in motor speech planning or articulation difficulties. In fact, distortions, which have been linked to apraxia of speech (Ogar et al., 2007; Duffy, 2013), were present to the same extent in both languages. The higher proportion of filled pauses in L1 could suggest a greater word finding problem in L1 compared to L2. However, LJ produced more nouns (as a proportion of narrative words) in L1 than in L2, while proportion of verbs was the same in L1 and L2. Furthermore, LJ used words of higher frequency in L2. This may indicate different levels of proficiency in L1 and L2. It must be noted here that lexical diversity was similar in both languages and that LJ made more semantic errors in L2. Greater number of filled pauses in L1 than in L2 may thus be explained with respect to the use of low frequency words and complex syntactic structures (Levelt, 1983; Ferreira et al., 1996), which is the case for the L1 narratives.

LJ produced a higher proportion of closed class words in L2 than in his L1 narratives. Nevertheless, this result must be interpreted by taking into account the increased rate of pronouns in L2. The proportion of pronouns was almost double in L2, but this can be explained by the underlying differences between French and Greek. As previously mentioned, Greek is a null subject language, whereas in French the inclusion of a subject is obligatory, and pronouns are commonly used to denote the subject in a sentence. Moreover, in the story retell task in L2, LJ was repeatedly using a double subject (both a noun and a pronoun as a subject), e.g., “The boy he was...,” “the mouse it went...,” The frequent use of subject doubling (double subject marking) may have inflated this measure.

In terms of discourse productivity, LJ produced longer narratives in L2 than in L1. However, proportion of narrative to total words was higher in L1 than in L2. This suggests that he was more efficient in getting his message across in L1 than in L2. Grammatical productivity was also better in L1. His sentences in Greek were more elaborate and complex, as indicated by the higher elaboration and embedding indexes in L1.

## GENERAL DISCUSSION

Summarizing the information in respect to language acquisition and use, LJ is a late bilingual speaker who acquired French in adulthood through formal instruction and a 7-year-long day-to-day exposure in a French language environment. He has

been using both Greek and French on a daily basis ever since, residing in a Greek-speaking country. Taking into account his wife's evaluation of level of proficiency in L1 and L2, and current exposure to both languages, Greek, LJ's first language, is his dominant language. Greek was designated as his more proficient language on the global measure of language proficiency and received a higher total score on task specific measures (11/30 in comparison to 10/30 for French). LJ has never attained fluency in reading and does not write in French. However, LJ was evaluated as being equally proficient in speaking in both languages. Language exposure to the two languages was rated as equal on the respective global measure, whilst, across different settings, language exposure to Greek (28/60) was higher than to French (21/60). Yet, the same extent of exposure to L1 and L2 was reported for interaction with his family. Even though there are skills in which LJ is equally competent in both languages and settings in which both languages are used at the same extent, taken together these results suggest that Greek is his dominant language. These results underlie the complexity of the bilingual experience and illustrate the difficulty in determining language dominance that has been attested by several researchers (Treffers-Daller, 2015).

In the present study, we predicted a similar pattern of impairment in both languages and a greater impairment in L2. Altogether, results suggest a slightly worse performance in LJ's second, non-dominant language for lexical and grammatical production and the presence of a similar pattern of impairment in both languages. Our predictions are therefore only partially supported.

According to Ullman (2001), L1 lexical processing is based on declarative memory, whereas syntactic and morphological processing on procedural memory. This is also the case for L2 when it is acquired at an early age. Given the fact that LJ is a late bilingual speaker, we would expect him to rely more on declarative memory for complex syntactic and morphological processing in L2 and on procedural memory processes for grammatical processing in L1. Increasing reliance on explicit processing for L2 could also be expected because French was learned formally (Paradis, 1994). Ullman (2001) has proposed that with extended practice and higher proficiency, L2 grammatical processing may increasingly rely on procedural memory.

However, a similar pattern of performance in L1 and L2 indicates that the same organizational principles underlie the two languages (Filley et al., 2006; Hernández et al., 2008; Druks and Weekes, 2013). In a late bilingual person with different levels of proficiency in L1 and L2, like LJ, similar patterns of impairment in both languages seem to indicate shared neural representations for the two languages. This conclusion is in line with the convergence hypothesis (Abutalebi and Green, 2007; Abutalebi, 2008) which posits that L1 and L2 depend on the same neural mechanisms and that L2 lexical and grammatical representations converge to L1 representations.

This model also predicts differences between L1 and L2, as late bilingual speakers need to recruit additional cognitive control resources to process their L2. Under this theoretical account, increased processing demands exist for LJ because French is

his non-dominant language. Differences between L1 and L2 may also be attributed to impaired control processes due to the underlying pathology of the nvfPPA. The executive deficit reported on neuropsychological assessment may account for the differences between the two languages. The cross-switching errors which were evident in the L2 personal narrative task support impairment in control functions. Cognitive control of L2 processing has been associated with the prefrontal cortex, the anterior cingulate cortex and the basal ganglia (Abutalebi and Green, 2007). Atrophy in the nvfPPA extends with disease progression into these regions, prefrontal cortex and anterior cingulate regions in particular (Grossman, 2010; Mesulam et al., 2014).

The fact that no significant differences were found between L1 and L2 seems to contradict our hypothesis. It must be noted however that long exposure to L2 and daily use of L2 at work and home may have played a role in preserving discourse abilities in L2. LJ uses and is exposed to French now for 36 years. Such a degree of exposure and use may play a determining role in L2 preservation. In fact, Abutalebi et al. (2015) found that differences between L1 and L2 suggesting an age of L2 acquisition effect are not present in elderly individuals. Nanchen et al. (2017) examining preservation of L1 and L2 in an immigrant population of late bilingual speakers with dementia, found that languages were equally preserved. They concluded that for elderly individuals, exposure and immersion are the main determinants of language preservation.

Our findings are consistent with a previous report (Zanini et al., 2011) of an early bilingual speaker with nvfPPA, where a decline in connected speech was found in both languages (Friulian and Italian), with the second language being impaired to a greater, but not to a significant degree. A qualitative similar pattern of deficits in L1 and L2 has been reported by Hernández et al. (2008) in an early, highly proficient Spanish-Catalan bilingual speaker with nvfPPA and Filley et al. (2006) in an early, proficient Chinese-English bilingual person with lvPPA. The only study which has investigated language abilities in a late bilingual speaker with nvfPPA was the study by Druks and Weekes (2013). Although grammatical production was not assessed, a parallel deterioration of lexical retrieval and grammatical knowledge in L1 (Hungarian) and L2 (English) was reported. This finding across two languages from different language families (Uralic and Indo-European, respectively) is similar to ours in that LJ was impaired, compared to controls, on both lexical and grammatical measures in his native language (Greek) and a parallel pattern of impairment was found in L2 (French), two structurally different languages albeit within the same family of languages.

In conclusion, we have found that LJ was impaired in lexical, discourse and grammatical productivity measures in his native language, Greek. A similar pattern of impairment was evident in his second language, French. Both L1 and L2 were affected to a similar degree. Lengthy exposure to L2 and regular activation of L2 through daily use may explain the preservation of discourse abilities in this non-dominant language. Connected speech analysis using QPA, fluency variables and error analysis has enabled the documentation of speech and language deficits

present in this case of the nfvPPA and the comparison of performance between the participant's languages.

A growing body of literature indicates that behavioral interventions in PPA can result in improvement of the targeted language function, although there are generalization and maintenance issues (Cadório et al., 2017). Research on bilingual aphasia rehabilitation after stroke has yielded inconsistent results regarding the pattern of cross-linguistic therapy effects (Goral and Conner, 2013). Evidence suggests that cross-language transfer of treatment gains is easier between two highly proficient languages, and from a less-proficient language to a more-proficient language (Ansaldi and Saidi, 2014). However, cross-language transfer also depends on factors such as postmorbid proficiency levels and linguistic similarity between languages (Goral et al., 2012). These data underline the clinical importance of determining language dominance and performance in both languages in bilinguals with PPA.

One limitation of the present study is the size of the speech samples. A minimum of 150 words has been suggested for QPA (Berndt et al., 2000). However, it was difficult to obtain samples of this size without extensive prompting. A second methodological limitation was the lack of control subjects. Ideally, neurotypical Greek-French bilingual individuals should have served as controls for this study. Furthermore, performance was assessed at one time point for both languages. Although we have data that show cognitive decline, we have not evaluated language performance at two time points. Thus, no conclusions can be drawn about the pattern of decline in each language and across languages. Finally, a factor that may have influenced results in L2 is the fact that LJ was assessed in both languages

by the same Greek-speaking clinician proficient in French. We know that healthy bilingual speakers' language choice is influenced by the social context and the linguistic background of the interlocutor (Blanco-Elorrieta and Pykkänen, 2017). Nevertheless, code-switching was observed only during the personal narrative in French. It could be a task related effect explained by LJ's difficulty in accessing the relevant words in French when talking about his daily job routine.

This study calls attention to factors such as language dominance, proficiency, patterns of use, and exposure to a language. These factors play a key role in assessing bilingual individuals with PPA and making clinical decisions based on the underlying linguistic and cognitive features.

## AUTHOR CONTRIBUTIONS

NK and MK designed the study. JP conducted the initial and follow-up assessments of the participant. Speech and language evaluation and QPA analysis was completed by NK. MK reviewed the QPA analysis. The manuscript was drafted by NK. Co-authors contributed to the final version of the manuscript.

## ACKNOWLEDGMENTS

We would like to thank LJ for participating in our study. We would also like to thank Maria Varkanitsa for the control group data, Eleni Dimakopoulou for her assistance with methodological and statistical analysis issues and Dr. Paraskevi Sakka, chairman of the Athens Alzheimer's Association, for encouraging and enabling research activities.

## REFERENCES

- Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychol.* 128, 466–478. doi: 10.1016/j.actpsy.2008.03.014
- Abutalebi, J., Canini, M., Della Rosa, P. A., Green, D. W., and Weekes, B. S. (2015). The neuroprotective effects of bilingualism upon the inferior parietal lobule: a structural neuroimaging study in aging Chinese bilinguals. *J. Neurolinguist.* 33, 3–13. doi: 10.1016/j.jneuroling.2014.09.008
- Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003
- Ansaldi, A. I., and Saidi, L. G. (2014). Aphasia therapy in the age of globalization: cross-linguistic therapy effects in bilingual aphasia. *Behav. Neurol.* 2014:603085. doi: 10.1155/2014/603085
- Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., et al. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology* 81, 329–336. doi: 10.1212/WNL.0b013e31829c5d0e
- Ash, S., McMillan, C., Gunawardena, D., Avants, B., Morgan, B., Khan, A., et al. (2010). Speech errors in progressive non-fluent aphasia. *Brain Lang.* 113, 13–20. doi: 10.1016/j.bandl.2009.12.001
- Ash, S., Moore, P., Vesely, L., Gunawardena, D., McMillan, C., Anderson, C., et al. (2009). Non-fluent speech in frontotemporal lobar degeneration. *J. Neurolinguist.* 22, 370–383. doi: 10.1016/j.jneuroling.2008.12.001
- Berndt, R., Wayland, S., Rochon, E., Saffran, E., and Schwartz, M. (2000). *Quantitative Production Analysis: A Training Manual for the Analysis of Aphasic Sentence Production*. Hove: Psychology Press.
- Blanco-Elorrieta, E., and Pykkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: the spatiotemporal dynamics of adaptive language control. *J. Neurosci.* 37, 9022–9036. doi: 10.1523/JNEUROSCI.0553-17.2017
- Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., and Cappa, S. F. (2017). Connected speech in neurodegenerative language disorders: a review. *Front. Psychol.* 8:269. doi: 10.3389/fpsyg.2017.00269
- Breining, B. L., Lala, T., Martínez Cuitiño, M., Manes, F., Peristeri, E., Tsapkini, K., et al. (2015). A brief assessment of object semantics in primary progressive aphasia. *Aphasiology* 29, 488–505. doi: 10.1080/02687038.2014.973360
- Bryant, L., Ferguson, A., and Spencer, E. (2016). Linguistic analysis of discourse in aphasia: a review of the literature. *Clin. Linguist. Phon.* 30, 489–518. doi: 10.3109/02699206.2016.1145740
- Cadório, I., Lousada, M., Martins, P., and Figueiredo, D. (2017). Generalization and maintenance of treatment gains in primary progressive aphasia (PPA): a systematic review. *Int. J. Lang. Commun. Disord.* 52, 543–560. doi: 10.1111/1460-6984.12310
- Cherney, L. R. (2010). Oral reading for language in aphasia: impact of aphasia severity on cross-modal outcomes in chronic nonfluent aphasia. *Semin. Speech Lang.* 31, 42–51. doi: 10.1055/s-0029-1244952
- Costa, A., Calabria, M., Marne, P., Hernández, M., Juncadella, M., Gascón-Bayarri, J., et al. (2012). On the parallel deterioration of lexico-semantic processes in the bilinguals' two languages: evidence from Alzheimer's disease. *Neuropsychologia* 50, 740–753. doi: 10.1016/j.neuropsychologia.2012.01.008
- Crawford, J. R., and Garthwaite, P. H. (2012). Single-case research in neuropsychology: a comparison of five forms of t-test for comparing a case to controls. *Cortex* 48, 1009–1016. doi: 10.1016/j.cortex.2011.06.021
- Davis, B. H., and MacLagan, M. (2009). Examining pauses in Alzheimer's discourse. *Am. J. Alzheimer's Dis. Other Dementias* 24, 141–154. doi: 10.1177/1533317508328138
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *Int. Rev. Appl. Linguist. Lang. Teach.* 54, 113–132. doi: 10.1515/iral-2016-9993



- De Jong, N. H., and Bosker, H. R. (2013). "Choosing a threshold for silent pauses to measure second language fluency," in *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech* (Stockholm), 17–20.
- Druks, J., and Weekes, B. S. (2013). Parallel deterioration to language processing in a bilingual speaker. *Cogn. Neuropsychol.* 30, 578–596. doi: 10.1080/02643294.2014.882814
- Duffy, J. R. (2013). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St. Louis, MO: Elsevier Mosby.
- Duffy, J. R., Strand, E. A., and Josephs, K. A. (2014). Motor speech disorders associated with primary progressive aphasia. *Aphasiology* 28, 1004–1017. doi: 10.1080/02687038.2013.869307
- Farooqi-Shah, Y., and Waked, A. N. (2010). Grammatical category dissociation in multilingual aphasia. *Cogn. Neuropsychol.* 27, 181–203. doi: 10.1080/02643294.2010.509340
- Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., and McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 324–335. doi: 10.1037/0278-7393.22.2.324
- Filley, C. M., Ramsberger, G., Menn, L., Wu, J., Reid, B. Y., and Reid, A. L. (2006). Primary progressive aphasia in a bilingual woman. *Neurocase* 12, 296–299. doi: 10.1080/13554790601126047
- Fountoulakis, K. N., Tsolaki, M., Chantzi, H., and Kazis, A. (2000). Mini mental state examination (MMSE): a validation study in Greece. *Am. J. Alzheimers Dis.* 15, 342–345. doi: 10.1177/153331750001500604
- Fountoulakis, K. N., Tsolaki, M., Iacovides, A., Yesavage, J., O'Hara, R., Kazis, A., et al. (1999). The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. *Aging Clin. Exp. Res.* 11, 367–372.
- Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., et al. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex* 55, 43–60. doi: 10.1016/j.cortex.2012.12.006
- Fraser, K. C., Rudzicz, F., Graham, N., and Rochon, E. (2013). "Automatic speech recognition in the diagnosis of primary progressive aphasia," in *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies* (Grenoble), 47–54.
- Fyndanis, V., Varlokosta, S., and Tsapkini, K. (2013). (Morpho)syntactic comprehension in agrammatic aphasia: evidence from Greek. *Aphasiology* 27, 398–419. doi: 10.1080/02687038.2013.770817
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, I., et al. (2012). *MAIN Multilingual Assessment Instrument for Narratives*. Berlin: ZAS Papers in Linguistics.
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balciuniene, I., et al. (2015). "Assessment of Narrative Abilities in Bilingual Children," in *Assessing Multilingual Children*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 243–269.
- Gollan, T. H., Salmon, D. P., Montoya, R. I., and da Pena, E. (2010). Accessibility of the nondominant language in picture naming: a counterintuitive effect of dementia on bilingual language production. *Neuropsychologia* 48, 1356–1366. doi: 10.1016/j.neuropsychologia.2009.12.038
- Goodglass, H., Kaplan, E., and Barresi, B. (2013). *Boston Diagnostic Aphasia Examination-Short Form in Greek Language*. Athens: Gotsis Publication.
- Goral, M., and Conner, P. S. (2013). Language disorders in multilingual and multicultural populations. *Annu. Rev. Appl. Linguist.* 33, 128–161. doi: 10.1017/S026719051300010X
- Goral, M., Rosas, J., Conner, P. S., Maul, K. K., and Obler, L. K. (2012). Effects of language proficiency and language of the environment on aphasia therapy in a multilingual. *J. Neurolinguist.* 25, 538–551. doi: 10.1016/j.jneuroling.2011.06.001
- Gordon, J. K. (2006). A quantitative production analysis of picture description. *Aphasiology* 20, 188–204. doi: 10.1080/02687030500472777
- Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014. doi: 10.1212/WNL.0b013e31821103e6
- Graham, N. L., Leonard, C., Tang-Wai, D. F., Black, S., Chow, T. W., Scott, C. J., et al. (2016). Lack of frank agrammatism in the nonfluent agrammatic variant of primary progressive aphasia. *Dement. Geriatr. Cogn. Dis. Extra* 6, 407–423. doi: 10.1159/000448944
- Graham, N. L., Patterson, K., and Hodges, J. R. (2004). When more yields less: speaking and writing deficits in nonfluent progressive aphasia. *Neurocase* 10, 141–155. doi: 10.1080/13554790409609945
- Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *J. Cogn. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377
- Grosjean, F. (1994). *Individual Bilingualism. The Encyclopedia of Language and Linguistics, 1656–1660*. Oxford: Pergamon Press.
- Grossman, M. (2010). Primary progressive aphasia: clinicopathological correlations. *Nat. Rev. Neurol.* 6, 88–97. doi: 10.1038/nrneurol.2009.216
- Hartsuiker, R. J., Pickering, M. J., and Velkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychol. Sci.* 15, 409–414. doi: 10.1111/j.0956-7976.2004.00693.x
- Hernández, M., Cano, A., Costa, A., Sebastián-Gallés, N., Juncadella, M., and Gascón-Bayarri, J. (2008). Grammatical category-specific deficits in bilingual aphasia. *Brain Lang.* 107, 68–80. doi: 10.1016/j.bandl.2008.01.006
- Higby, E., Kim, J., and Obler, L. K. (2013). Multilingualism and the Brain. *Annu. Rev. Appl. Linguist.* 33, 68–101. doi: 10.1017/S0267190513000081
- Hillis, A. E., Heidler-Gary, J., Newhart, M., Chang, S., Ken, L., and Bak, T. H. (2006). Naming and comprehension in primary progressive aphasia: the influence of grammatical word class. *Aphasiology* 20, 246–256. doi: 10.1080/02687030500473262
- Ivanova, I., Salmon, D. P., and Gollan, T. H. (2014). Which language declines more? Longitudinal versus cross-sectional decline of picture naming in bilinguals with Alzheimer's disease. *J. Int. Neuropsychol. Soc.* 20, 534–546. doi: 10.1017/S1355617714000228
- Kambanaros, M. (2009). Group effects of instrumentality and name relation on action naming in bilingual anomic aphasia. *Brain Lang.* 110, 29–37. doi: 10.1016/j.bandl.2009.01.004
- Kambanaros, M. (2010). Action and object naming versus verb and noun retrieval in connected speech: comparisons in late bilingual Greek-English anomic speakers. *Aphasiology* 24, 210–230. doi: 10.1080/02687030902958332
- Kambanaros, M. (2016). Verb and noun word retrieval in bilingual aphasia: a case study of language-and modality-specific levels of breakdown. *Int. J. Biling. Educ. Biling.* 19, 169–184. doi: 10.1080/13670050.2015.1037717
- Kambanaros, M., and Grohmann, K. K. (2012). Bilingual primary progressive aphasia for Greek, English, and Czech. *J. Neurolinguist.* 25, 520–537. doi: 10.1016/j.jneuroling.2011.01.006
- Kambanaros, M., and van Steenbrugge, W. (2006). Noun and verb processing in Greek-English bilingual individuals with anomic aphasia and the effect of instrumentality and verb-noun name relation. *Brain Lang.* 97, 162–177. doi: 10.1016/j.bandl.2005.10.001
- Kavé, G., Leonard, C., Cupit, J., and Rochon, E. (2007). Structurally well-formed narrative production in the face of severe conceptual deterioration: a longitudinal case study of a woman with semantic dementia. *J. Neurolinguist.* 20, 161–177. doi: 10.1016/j.jneuroling.2006.06.003
- Khachatryan, E., Vanhoof, G., Beyens, H., Goeleven, A., Thijs, V., and Van Hulle, M. M. (2016). Language processing in bilingual aphasia: a new insight into the problem. *Wiley Interdiscipl. Rev. Cogn. Sci.* 7, 180–196. doi: 10.1002/wcs.1384
- Knibb, J. A., Woollams, A. M., Hodges, J. R., and Patterson, K. (2009). Making sense of progressive non-fluent aphasia: an analysis of conversational speech. *Brain* 132, 2734–2746. doi: 10.1093/brain/awp207
- Konstantinopoulou, E., Kosmidis, M. H., Ioannidis, P., Kiosseoglou, G., Karacostas, D., and Taskos, N. (2011). Adaptation of Addenbrooke's cognitive examination – revised for the Greek population. *Eur. J. Neurol.* 18, 442–447. doi: 10.1111/j.1468-1331.2010.03173.x
- Krivokapi, J. (2007). Prosodic planning: effects of phrasal length and complexity on pause duration. *J. Phon.* 35, 162–179. doi: 10.1016/j.wocn.2006.04.001
- Larner, A. J. (2012). Progressive non-fluent aphasia in a bilingual subject: Relative preservation of "Mother Tongue." *J. Neuropsychiatry Clin. Neurosci.* 24, E9–E10. doi: 10.1176/appi.neuropsych.11010019
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition* 14, 41–104. doi: 10.1016/0010-0277(83)90026-4
- Leyton, C. E., and Ballard, K. J. (2016). Primary progressive aphasia: conceptual evolution and challenges. *Neurosci. Neuroeconom.* 5, 9–18. doi: 10.2147/NAN.S102848



- Liu, H., and Cao, F. (2016). L1 and L2 processing in the bilingual brain: a meta-analysis of neuroimaging studies. *Brain Lang.* 159, 60–73. doi: 10.1016/j.bandl.2016.05.013
- Lorenzen, B., and Murray, L. L. (2008). Bilingual aphasia: a theoretical and clinical review. *Am. J. Speech Lang. Pathol.* 17, 299–317. doi: 10.1044/1058-0360(2008/026)
- Machado, Á., Rodrigues, M., Simões, S., Santana, I., and Soares-Fernandes, J. (2010). The Portuguese who could no longer speak French: primary progressive aphasia in a bilingual man. *J. Neuropsychiatry Clin. Neurosci.* 22, 123.e31–123.e32. doi: 10.1176/jnp.2010.22.1.123.e31
- Mack, J. E., Chandler, S. D., Meltzer-Asscher, A., Rogalski, E., Weintraub, S., et al. (2015). What do pauses in narrative production reveal about the nature of word retrieval deficits in PPA? *Neuropsychologia* 77, 211–222. doi: 10.1016/j.neuropsychologia.2015.08.019
- Manchon, M., Buetler, K., Colombo, F., Spierer, L., Assal, F., and Annoni, J.-M. (2015). Impairment of both languages in late bilinguals with dementia of the Alzheimer type. *Bilingual. Lang. Cogn.* 18, 90–100. doi: 10.1017/S1366728914000194
- Marcotte, K., Graham, N. L., Fraser, K. C., Meltzer, J. A., Tang-Wai, D. F., Chow, T. W., et al. (2017). White matter disruption and connected speech in non-fluent and semantic variants of primary progressive aphasia. *Dement. Geriatr. Cogn. Dis. Extra* 7, 52–73. doi: 10.1159/000456710
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007/067)
- Marini, A., Andreetta, S., Del Tin, S., and Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology* 25, 1372–1392. doi: 10.1080/02687038.2011.584690
- Mendez, M. F., Perryman, K. M., Ponton, M. O., and Cummings, J. L. (1999). Bilingualism and dementia. *J. Neuropsychiatry Clin. Neurosci.* 11, 411–412. doi: 10.1176/jnp.11.3.411
- Mesulam, M.-M., Rogalski, E. J., Wieneke, C., Hurley, R. S., Geula, C., Bigio, E. H., et al. (2014). Primary progressive aphasia and the evolving neurology of the language network. *Nat. Rev. Neurol.* 10, 554–569. doi: 10.1038/nrneurol.2014.159
- Meteyard, L., and Patterson, K. (2009). The relation between content and structure in language production: an analysis of speech errors in semantic dementia. *Brain Lang.* 110, 121–134. doi: 10.1016/j.bandl.2009.03.007
- Nanchen, G., Abutalebi, J., Assal, F., Manchon, M., Démonet, J. F., and Annoni, J. M. (2017). Second language performances in elderly bilinguals and individuals with dementia: the role of L2 immersion. *J. Neurolinguistics* 43, 49–58. doi: 10.1016/j.jneuroling.2016.09.004
- New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychol.* 101, 447–462. doi: 10.3406/psy.2001.1341
- Ogar, J. M., Dronkers, N. F., Brambati, S. M., Miller, B. L., and Gorno-Tempini, M. L. (2007). Progressive nonfluent aphasia and its characteristic motor speech deficits. *Alzheimer Dis. Associated Disord.* 21, S23–S30. doi: 10.1097/WAD.0b013e31815d19fe
- Paradis, M. (1994). “Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism,” in *Implicit and Explicit Learning of Second Languages*, ed N. Ellis (London: Academic Press), 393–419.
- Paradis, M. (2008). Bilingualism and neuropsychiatric disorders. *J. Neurolinguist.* 21, 199–230. doi: 10.1016/j.jneuroling.2007.09.002
- Patterson, K., Graham, N., Ralph, M. A. L., and Hodges, J. (2006). Progressive non-fluent aphasia is not a progressive form of non-fluent (post-stroke) aphasia. *Aphasiology* 20, 1018–1034. doi: 10.1080/02687030600739463
- Prévost, P. (2009). *The Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*. Philadelphia; Amsterdam: John Benjamins Publishing. doi: 10.1075/lald.51
- Protopapas, A., Tzakosta, M., Chalamandaris, A., and Tsiakoulis, P. (2012). IPLR: An online resource for Greek word-level and sublexical information. *Lang. Resources Evalu.* 46, 449–459. doi: 10.1007/s10579-010-9130-z
- Roberts, I., and Holmberg, A. (2010). “Introduction: parameters in minimalist theory,” in *Parametric Variation: Null Subjects in Minimalist Theory*, ed T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan (Cambridge: Cambridge University Press), 1–57.
- Rochon, E., Saffran, E. M., Berndt, R. S., and Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: further development and new data. *Brain Lang.* 72, 193–218. doi: 10.1006/brln.1999.2285
- Rogalski, E., Cobia, D., Harrison, T. M., Wieneke, C., Thompson, C. K., Weintraub, S., et al. (2011). Anatomy of language impairments in primary progressive aphasia. *J. Neurosci.* 31, 3344–3350. doi: 10.1523/JNEUROSCI.5544-10.2011
- Saffran, E. M., Berndt, R. S., and Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: procedure and data. *Brain Lang.* 37, 440–479. doi: 10.1016/0093-934X(89)90030-8
- Sajjadi, S. A., Patterson, K., Tomek, M., and Nestor, P. J. (2012). Abnormalities of connected speech in the non-semantic variants of primary progressive aphasia. *Aphasiology* 26, 1219–1237. doi: 10.1080/02687038.2012.710318
- Salvatierra, J., Rosselli, M., Acevedo, A., and Duara, R. (2007). Verbal fluency in bilingual Spanish/English Alzheimer's disease patients. *Am. J. Alzheimers Dis. Other Dement.* 22, 190–201. doi: 10.1177/1533317507301792
- Santos-Santos, M. A., Mandelli, M. L., Binney, R. J., Ogar, J., Wilson, S. M., Henry, M. L., et al. (2016). Cross-sectional and longitudinal features of non-fluent/agrammatic primary progressive aphasia with underlying corticobasal degeneration or progressive supranuclear palsy pathology. *JAMA Neurol.* 73, 733–742. doi: 10.1001/jamaneurol.2016.0412
- Sapolsky, D., Bakkour, A., Negreira, A., Nalipinski, P., Weintraub, S., Mesulam, M. M., et al. (2010). Cortical neuroanatomic correlates of symptom severity in primary progressive aphasia. *Neurology* 75, 358–366. doi: 10.1212/WNL.0b013e3181ea15e8
- Sebastian, R., Laird, A. R., and Kiran, S. (2011). Meta-analysis of the neural representation of first language and second language. *Appl. Psycholinguist.* 32, 799–819. doi: 10.1017/S0142716411000075
- Simos, P. G., Kasselimis, D., and Mouzaki, A. (2011). Effects of demographic variables and health status on brief vocabulary measures in Greek. *Aphasiology* 25, 492–504. doi: 10.1080/02687038.2010.512120
- Simos, P. G., Sideridis, G. D., Kasselimis, D., and Mouzaki, A. (2013). Reading fluency estimates of current intellectual function: demographic factors and effects of type of stimuli. *J. Int. Neuropsychol. Soc.* 19, 355–361. doi: 10.1017/S1355617712001518
- Sloetjes, H., and Wittenburg, P. (2008). “Annotation by category – ELAN and ISO DCR,” in *Proceedings of the 6th International Conference on Language Resources and Evaluation* (Marrakech), 816–820. Available online at: <https://tla.mpi.nl/tools/tla-tools/elan/>
- Thompson, C. K., Cho, S., Hsu, C. J., Wieneke, C., Rademaker, A., Weitner, B. B., et al. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology* 26, 20–43. doi: 10.1080/02687038.2011.584691
- Treffers-Daller, J. (2015). “The construct of language dominance, its operationalization and measurement,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalan and J. Treffers-Daller (Cambridge: Cambridge University Press), 235–265.
- Tsimpli, I., Sorace, A., Heycock, C., and Filiaci, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *Int. J. Bilingual.* 8, 257–277. doi: 10.1177/13670069040080030601
- Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nat. Rev. Neurosci.* 2, 717–726. doi: 10.1038/35094573
- Van Hout, R., and Vermeer, A. (2007). “Comparing measures of lexical richness,” in *Modelling and Assessing Vocabulary Knowledge*, eds H. Daller, J. Milton, and J. Treffers-Daller (Cambridge: Cambridge University Press), 93–115.
- Varkanitsa, M. (2012). “Quantitative and error analysis of connected speech: evidence from Greek-speaking patients with aphasia and normal speakers,” in *Current Trends in Greek Linguistics*, eds G. Fragaki, T. Georgakopoulos, and C. Themistocleous (Newcastle: Cambridge Scholars Publishing), 313–338.
- Weber, K., and Indefrey, P. (2009). Syntactic priming in German–English bilinguals during sentence comprehension. *Neuroimage* 46, 1164–1172. doi: 10.1016/j.neuroimage.2009.03.040
- Weber, K., Luther, L., Indefrey, P., and Hagoort, P. (2016). Overlap and differences in brain networks underlying the processing of complex sentence structures in second language users compared with native speakers. *Brain Connect.* 6, 345–355. doi: 10.1089/brain.2015.0383

- Wertz, R. T., LaPointe, L. L., and Rosenbek, J. C. (1984). *Apraxia of Speech: The Disorder and its Management*. New York, NY: Grune and Stratton.
- Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., et al. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain* 133, 2069–2088. doi: 10.1093/brain/awq129
- Youmans, G., Holland, A., Muñoz, M., and Bourgeois, M. (2005). Script training and automaticity in two individuals with aphasia. *Aphasiology* 19, 435–450. doi: 10.1080/02687030444000877
- Zanini, S., Angeli, V., and Tavano, A. (2011). Primary progressive aphasia in a bilingual speaker: a single-case study. *Clin. Linguist. Phon.* 25, 553–564. doi: 10.3109/02699206.2011.566464

**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 Karpathiou, Papatriantafyllou and Kambanaros. 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.



# First Language Attrition and Dominance: Same Same or Different?

Barbara Köpke<sup>1\*</sup> and Dobrinka Genevaska-Hanke<sup>2\*</sup>

<sup>1</sup> Octogone-Lordat (Interdisciplinary Research Unit), University of Toulouse, Toulouse, France, <sup>2</sup> Department of English, University of Oldenburg, Oldenburg, Germany

## OPEN ACCESS

### Edited by:

Carlos Acuña-Fariña,  
Universidade de Santiago  
de Compostela, Spain

### Reviewed by:

Arturo Hernandez,  
University of Houston, United States  
Cristina Maria Flores,  
University of Minho, Portugal

### \*Correspondence:

Barbara Köpke  
bkopke@univ-tlse2.fr  
Dobrinka Genevaska-Hanke  
dobrinka.genevaska.hanke@uni-  
oldenburg.de

### Specialty section:

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

**Received:** 15 May 2018

**Accepted:** 24 September 2018

**Published:** 06 November 2018

### Citation:

Köpke B and Genevaska-Hanke D  
(2018) First Language Attrition and  
Dominance: Same Same or  
Different?. *Front. Psychol.* 9:1963.  
doi: 10.3389/fpsyg.2018.01963

We explore the relationship between first language attrition and language dominance, defined here as the relative availability of each of a bilingual's languages with respect to language processing. We assume that both processes might represent two stages of one and the same phenomenon (Schmid and Köpke, 2017; Köpke, 2018). While many researchers agree that language dominance changes repeatedly over the lifespan (e.g., Silva-Corvalan and Treffers-Daller, 2015), little is known about the precise time scales involved in dominance shifts and attrition. We investigate these time scales in a longitudinal case study of pronominal subject production by a near-native L2-German (semi-null subject and topic-drop but non-pro-drop) and L1-Bulgarian (pro-drop) bilingual speaker with 17 years of residence in Germany. This speaker's spontaneous speech showed a significantly higher rate of overt pronominal subjects in her L1 than the controls' rates when tested in Germany. After 3 weeks of L1-reexposure in Bulgaria, however, attrition effects disappeared and the overt subject rate fell within the monolinguals' range (Genevaska-Hanke, 2017). The findings of this first investigation are now compared to those of a second investigation 5 years later, involving data collection in both countries with the result that after 17 years of immigration, no further attrition was attested and the production of overt subjects remained monolingual-like for the data collections in both language environments. The discussion focuses on the factors that are likely to explain these results. First, these show that attrition and language dominance are highly dependent on immediate language use context and change rapidly when the language environment is modified. Additionally, the data obtained after L1-reexposure illustrate that time scales involved in dominance shift or attrition are much shorter than previously thought. Second, the role of age of acquisition in attrition has repeatedly been acknowledged. The present study demonstrates that attrition of a highly entrenched L1 is a phenomenon affecting language processing only temporarily and that it is likely to regress quickly after reexposure or return to balanced L1-use. The discussion suggests that dominance shift and attrition probably involve similar mechanisms and are influenced by the same external factors, showing that both may be different steps of the same process.

**Keywords:** bilingualism, attrition, dominance, reexposure, time scales, stability, context dependence, null subjects

## INTRODUCTION

Over the last decades, research on language attrition has progressively become part of the field of bilingual development, together with studies on first language development, second language acquisition and age related changes in language use and/or cognition (see for instance, the chapters in De Bot and Schrauf, 2009). In such a perspective, attrition is often defined as “(...) the loss of language proficiency within an individual over time” (De Bot and Schrauf, 2009, p. 11). In many studies on attrition, researchers seem to take for granted that attrition involves or may even be causally linked to a change in language dominance, which we refer to as *dominance shift* in the following. In the case of attrition of the first language (L1), it is assumed that continuous immersion in a second language (L2) environment will lead to a growing influence of the L2 on the L1, which is then becoming the non-dominant language. In the case of L2 attrition, an individual who was previously immersed in an L2 environment returns to the L1 environment, where the L1 is regaining dominance again (e.g., Hansen, 1999).

The interest in language dominance and in the factors involved in it has considerably increased in recent years (e.g., this issue; Silva-Corvalan and Treffers-Daller, 2015). This is fortunate since previous research often lacked precision with respect to what was meant by language dominance. Furthermore, the link between language dominance and attrition on the one hand, and these two processes and cross-linguistic influence on the other hand, are not understood very clearly. Already in 2004 Köpke and Schmid suggested a relationship between attrition and dominance and hypothesized that “... even if a reversal in language dominance is not necessarily followed by attrition, it is most likely that attrition is preceded by such a reversal ...” (2004, p. 12). In the same vein, these authors proposed recently that L1 attrition may “...refer to any of the phenomena that arise in the native language of a sequential bilingual as the consequence of the co-activation of language, cross-linguistic transfer or disuse” (Schmid and Köpke, 2017, p. 637), suggesting a similarity between different processes of interaction between the languages of a bilingual. Such a suggestion is not incompatible with recent conceptions of language dominance. For instance, it has been proposed that language dominance is not a uni-dimensional phenomenon but a complex construct involving a variety of dimensions and remaining relatively independent for different linguistic domains (Birdsong, 2018). This is undoubtedly also the case for attrition. However, we think that it is probably premature at the present stage to conclude that language dominance and attrition refer to one and the same process. In order to examine this question, more data on bilingual development at different points in time during the life of an individual are needed in order to investigate the linguistic changes observable at different time-scales – days, weeks, or years – after modifications in the linguistic environment (including loss of language contact and subsequent reexposure), or other factors (such as attitude changes) that are still poorly understood at the present moment.

With this work, we aim to contribute to a better understanding of the links between dominance shift and L1 attrition. In order to do so, we will first provide a short overview of different possible definitions and operationalizations of the concept of language dominance. We will then focus on the dynamics of changes in language dominance and attrition, through a review of studies focussing on the time scales of these processes in longitudinal studies. Special attention will be paid to the issue of reexposure to a previously attrited (or supposedly attrited) language, a question that has not received much attention until now despite its potential interest for a more comprehensive understanding of the dynamics of dominance shift and attrition as well as the factors that may influence them. We then present the findings from a longitudinal study on subject use in a Bulgarian-German late bilingual tested at four investigation times over a period of 5 years. While data obtained with single case studies are rather limited and generally disallow for generalizations, they represent the type of data crucially needed. Data of this kind lead us to a discussion of the external factors that may explain the effects of reexposure observed and on their relevance for the debate on the links between dominance shift and attrition.

## LANGUAGE DOMINANCE IN RESEARCH WITH BILINGUAL SPEAKERS

While many studies on bilingualism refer to the concept of language dominance in the description of their participants or discussion of their results, the term itself is, to our knowledge, most of the time not clearly defined. What appears in the use made of the term seems to refer to quite different conceptions of language dominance and further depends on whether the studies focus on bilingual children or adults.

A lot of studies implicitly or explicitly define *dominance* as the relative proficiency in each of the languages of a bilingual (e.g., Silva-Corvalan and Treffers-Daller, 2015). In such a perspective, studies focusing on children most of the time refer to a *strong* and a *weak* language established through production measures such as mean length of utterance (MLU), vocabulary size, or overall number of utterances (De Houwer, 2009). Language dominance in children has furthermore been related to rates of mixing and directionality of cross-linguistic interference (see Unsworth, 2015, for a review). Other studies establish language dominance mainly through the language of the environment. In heritage language research, for instance, many researchers employ the term *dominant language* to refer to the majority language (e.g., Rothman, 2009). Others refer to exposure criteria for each child, as Mayr et al. (2014) who talk about *English-only-homes* vs. *Welsh-only-homes*.

Recently, Unsworth (2015) demonstrated that proficiency and exposure criteria are closely linked in young children in an investigation of 18 Dutch–English bilingual children, aged 2–4 years. Such studies seem to substantiate the claim that exposure is a valid indicator of dominance (in terms of relative proficiency in each language) and has led the author to suggest that exposure can be used as a proxy of language dominance. Others go even further and propose that language dominance is a complex factor,



involving proficiency-related components as well as both external (input) and functional (use, context) components (Montrul, 2015; see also de Almeida et al., 2017; Hamann and Abed Ibrahim, 2017). But on the whole, for the developing languages of a bilingual child, relative proficiency in each language seems to be the principal criterion of language dominance adopted in current research.

With respect to adult bilinguals, Wei (2007) referred to a mixture of proficiency and exposure criteria when he proposed that a *dominant language* is the one the bilingual is more proficient in and the one that is used more frequently. However, while for bilingual children the links between frequency of use or input and proficiency are evident in most studies, this is much less straightforward when adult bilinguals are considered. In order to compensate for the absence of such a direct link between use and proficiency in adults, many studies seem to seek to establish linguistic markers of dominance with a large variety of means for establishing relative proficiency (see Flege et al., 2002, for a summary). These include measures of processes involved in utterance planning and lexical access or directionality of code-switching/transfer (Daller, 2011), lexical richness (Treffers-Daller, 2011), discourse patterns (Flecken, 2011), fluency measures, and C-Tests (Daller et al., 2011), among others. The underlying rationale of these studies is similar to what is proposed in studies on children: balanced bilinguals will have similar proficiency measures with respect to various aspects of language use, while speakers who are dominant in one language will achieve higher scores and proficiency measures in that particular language. Thus, in these approaches, *dominance* equals increased proficiency.

However, some authors disagree with the point of view that dominance is mainly an issue of relative proficiency. For instance, Gertken et al. (2014) propose that dominance is independent from proficiency and that it is possible for a speaker to be dominant in a less proficient language. This is in line with a more psycholinguistic definition of dominance, based on the relative availability of each of the languages of a bilingual, as known from studies on lexical retrieval and access.

In a very early study, Lachman and Mistler-Lachman (1976) pursued the question whether an L2 could become dominant over an L1. With reference to models of information processing, they relate language dominance to “the ability to process a language” (p. 282, our translation), the dominant language being the one that is more easily processed. In other terms, in their study focussing on lexical retrieval of single words, they considered the dominant language to be “the language in which the person will retrieve words easier.” They furthermore distinguish the dominant language from what they call the *usual language* which is the language used predominantly. Contrary to most approaches to language dominance, they consider the increased use of the usual language as a necessary but insufficient condition for the establishment of language dominance. Hence, in their study, specific attention was paid to the selection criteria for the participants – they had to have highly predominant use of their L2 (i.e., more than 80% of the language use reported) in order to state that the L2 was the usual language. The authors then investigated whether the usual L2 was also the dominant language

of the participants. Six German–English late bilinguals were tested with a timed picture naming task, which was remarkably well-controlled in L1 and L2. Participants were aged between 22 and 48 years and had spoken predominantly English for a period of one to 27 years. The results showed that the participants who had spent less than 7 years in an L2 environment were slower to name pictures in L2 than in L1, while those who had spent 15 or more years<sup>1</sup> in an L2 environment showed the reverse pattern and were slower to name pictures in L1. While the time scales observed have to be considered with much caution given the limited number of participants, it is worth noting that this study was the very first empirical investigation of L1 attrition reported in the literature (although the term *attrition* was not used) and that the starting point of the authors was to look for a reversal in language dominance patterns.

Similar findings were reported a couple of years later with respect to lexical access in a lexical decision paradigm. Frenck-Mestre (1993) investigated lexical recognition in the L2 of 20 Anglophone undergraduate students attending a French university at the time of testing and showed that skilled bilinguals who had been living in an L2 environment for 3 years or longer, responded faster to L2 words than to L1 words, while beginning bilinguals who had been living in an L2 environment for less than 6 months responded faster in L1. This observation was interpreted as a shift in language dominance illustrating the role of previous experience and actual contact with the language in a word identification experiment.

The findings of these studies suggest an alternative interpretation of the concept of language dominance in terms of *processing facility* or *processing ease* (a label used very recently by Birdsong, 2018, p. 2). In this view, a dominance shift arises as a consequence of increased use of an L2 and leads to a delay in L1 processing. More precisely, the findings of these early studies suggest, in a very preliminary way, that a dominance shift may arise more quickly in perceptual processing, involved in word recognition tasks (3 years in Frenck-Mestre, 1993) than in language production as evidenced by results of the naming task used by Lachman and Mistler-Lachman (1976). Similar observations have been reported by Mägiste (1979) in another early study with 163 bilingual adolescents who showed shorter processing times for L2 after 4–6 years of L2 immersion in comprehension tasks and after 6 years in production tasks (see Köpke and Schmid, 2004, for more details). Moreover, this view assumes that a language (generally the language of the environment) may be more accessible for psycholinguistic processing even though its production does not always equate high-proficiency with respect to phonological, grammatical and even other lexical features (as established with proficiency measures). Such a view is easily implemented in the context of lexical processing and corroborated by recent studies showing that even a few months of immersion in a foreign language, as is typically the case for students in a study abroad program, may lead to increased response times in L1 picture naming (Baus et al., 2013) or to a reduction in lexical retrieval in a verbal fluency task (Linck et al., 2009). Whether processing ease may be

<sup>1</sup>But note that there were no participants in between these two values.

at play in a similar way with respect to syntax is less clear yet, but there are studies suggesting that preferences in syntactic online processing may, similarly, be influenced by language context, e.g., with respect to relative clause attachment (Dussias, 2004; Dussias and Sagarra, 2007).

In general, such insights from studies on dominance shift have not sufficiently been taken into account in attrition research – despite the fact that they are perfectly compatible with frequency-based accounts commonly referred to in attrition research, such as the Activation Threshold Hypothesis, ATH (Paradis, 2004; Paradis M., 2007), which predicts that the availability of linguistic material in the bilingual mind will be dependent on frequency and recency of use. However, L1 as well as L2 attrition studies have suggested that attrition cannot be explained by frequency of use alone (see Schmid, 2007 and Mehotcheva and Köpke, 2019, for reviews). Instead, it has been proposed that only a combination of factors may provide the conditions for attrition to arise (see also Schmid and Yilmaz, 2018). Similar perspectives have been taken up with respect to language dominance. It has been proposed that the dominant language is not only the more active language in bilingual processing (and the one related to automaticity), but that it is also influenced by extralinguistic factors such as language attitude for instance (for a review see Gertken et al., 2014). These authors further suggest that dominance may be domain-specific in an individual. This corroborates the idea that bilingual dominance is a complex concept, arising from a combination of criteria (Birdsong, 2014, but also Grosjean, 1998; Flege et al., 2002). Recently, a number of test tools have been proposed that take into account a complexity of this kind. For instance, Dunn and Fox Tree (2009) base their short gradient dominance scale on three main criteria: percentage of use of each language, age of acquisition and age of comfort for both languages. The scale further involves a short question on restructuring of language fluency due to changes in the environment. Gertken et al. (2014) propose a more detailed questionnaire that focuses on language history, use, proficiency and attitudes.

What becomes evident in this review is that dominance shift and attrition are established with similar measures and seem to be influenced by the same factors. Adopting a psycholinguistic approach, it is not unlikely that both processes rely on very similar mechanisms and perhaps represent different stages of a continuum. Following the distinction made by Lachman and Mistler-Lachman (1976), the usual language (as established through frequency of use) will at some point of time become dominant (more readily available for language processing). Whether a dominance change of this kind is equivalent to attrition or whether attrition arises at a later stage of the process, is not clear yet. With respect to L1 attrition in adults, which is the type of attrition the current study focuses on, researchers seem to adopt one or the other standpoint depending heavily on their definition of attrition: if attrition is mainly seen as a phenomenon of on-line processing, dominance and attrition are identical (e.g., Schmid and Köpke, 2017; Schmid and Yilmaz, 2018); if attrition is defined as the restructuring of linguistic representations (e.g., Gürel, 2017; Tsimpli, 2017), then dominance change and attrition are likely to be different and arise at different stages of bilingual

development<sup>2</sup>. Köpke (2018) has recently proposed that we may talk about attrition when the processing of the non-dominant language is becoming so cumbersome that disfluencies may be perceived, but there is not much data on perceived attrition at the present moment (be it by the bilingual herself, or by other speakers). So for now, in order to better understand the link between dominance change and attrition, it is probably safer to increase the body of research on the time-scales involved in both processes.

## TIME SCALES OF DOMINANCE SHIFT AND ATTRITION

In order to investigate the temporariness of dominance shift and attrition processes, we need a clearer picture of the evolution of dominance along the lifespan, reflecting the modifications in exposure and use that arise in a bilingual life. However, to date, the tools that have been developed to establish language dominance do not allow us to capture multiple evolutions across the life-span, despite the efforts that have been made. The short scale proposed by Dunn and Fox Tree (2009) overvalues the acquisition context and attributes a lot of weight to the first acquired language and the question of a possible accent. While the possibility of a loss of fluency is taken up, the single question on this only allows for a binary response. The Bilingual Language Profile by Gertken et al. (2014) is much more detailed with respect to language background and present language use but it doesn't allow for the consideration of multiple changes in language use either. Thus, none of these tools allows for a satisfactory assessment of multiple dominance shifts and effects of reexposure as they may arise in attrition contexts.

In attrition research, despite consensus about the importance of time, operationalized as Length of Residence (LoR), surprisingly little is known about the time scales involved. Following the rationales of theoretical frameworks as the ATH and memory decay theories, many authors have assumed that L1 attrition in adult speakers is a slow process (Hutz, 2004). In addition, most empirical studies involve participants that have spent at least a decade in their new language environment, as was suggested in the settings of first language attrition studies, namely by Seliger and Vago (1991). However, these studies fail to provide evidence for any direct links between LoR and attrition. In most of them, the observed attrition effects are attributable to a small number of immigrants and are probably due to a complex interaction of multiple factors (e.g., Cherciov, 2013; Opitz, 2013). The number of longitudinal studies providing data on the evolution of attrition over time is still very limited and the conclusions of these studies invite us to revisit the concept of time in relation to attrition. The only group study among them (De Bot and Clyne, 1989, 1994) focused on 40 Dutch immigrants in Australia, who were re-examined 16 years after a first investigation of 200 participants in the early 70s (Clyne,

<sup>2</sup>Please note that this is very close to the continuum between performance and competence in attrition, proposed by Sharwood Smith (1983, p. 51) in the very beginnings of attrition research.

1981). While the first study suggested that elderly immigrants may suffer from L2 attrition after retirement and reinforce their L1, known as the *language reversion hypothesis*, the second study did not confirm any further changes in these immigrants, neither in L1 nor in L2. This result has been interpreted as evidence for the existence of some kind of a threshold in L2 (and L1?) knowledge, after which the language is no longer sensitive to further changes in use or exposure.

All other longitudinal studies are case studies. Ecke and Hall (2013) report on a study of tip of the tongue (TOT) states in a multilingual subject (five languages) who kept a diary about his TOT states during a period of 10 years. The study focused primarily on the interactions between his most frequently used L3 and L4 (English and Spanish) and his L1 German that was viewed as attriting due to reduced use throughout the study. The results documented the directionality of interference from the more dominant languages to the less dominant L1 and suggest that despite overall high resistance of the L1 to attrition there was a temporary impairment of the L1 in the initial stages of L3 and L4 immersion, where “the overall set of language systems comes out of balance” (p. 1). However, after this period, the L1 gained stability again, suggesting the temporary nature of the phenomenon. Two more studies examined the written production of long-term immigrants (Jaspaert and Kroon, 1992; Hutz, 2004). What all these studies suggest is that the most important changes take place during the first decade of immigration and that longitudinal data, collected after several decades of immersion in a second language environment, do not provide evidence for additional attrition. Given these results, it is surprising that many studies of attrition have continued to focus on immigrants with an LoR of more than 10 or even 15 years.

Some recent studies of L1 processing either in active bilinguals or in second language learners provide more data. For instance, Chang (2012) showed that native speakers of English learning Korean through an intensive language programme provide evidence for temporal changes at segmental, suprasegmental and global levels of pronunciation of their L1 and that this holds even for beginners. In the syntactic domain, Dussias and collaborators showed that previous exposure to specific sentence types may influence relative clause attachment in Spanish-English bilinguals (Dussias et al., 2014; see Schmid and Köpke, 2017, for a more detailed review). These studies suggest that immediate language context may influence language processing and lead to cross-linguistic influence as well as dominance effects on much smaller time scales than previously thought. Again, whether this equals attrition remains an open question. Most importantly, such insights raise the question of the effects of reexposure on the attrited language or a language that has become non-dominant.

However, the question of reexposure is still largely neglected in the field of research on attrition and bilingual development. A small number of studies conducted with adoptees, re-exposed to their native language later in life, mainly focus on the reminiscents of a childhood language and on the benefits of later relearning (e.g., Oh et al., 2019; Pierce et al., 2019). The only longitudinal study in the context of attrition we know of is a study describing dominance shifts in an English-Bulgarian bilingual child (Slavkov, 2015), but given the young age of the

child (1;7–2;3) it seems difficult to generalize to what happens in adults.

As far as adult immigrants are concerned, some evidence about reexposure is provided by Stolberg and Münch (2010) who examined a very long term immigrant with no L1 contact at all for 50 years. For the purpose of the study, the participant was interviewed every 2–3 months during a period of 4 years. The data shows a decrease of disfluencies and grammatical errors over time, suggesting that even such a reduced amount of language contact, as in the context of this study, may be sufficient to reactivate a first language. In the phonetic domain, adaptation to VOT standards of one or the other language in bilingual speakers has been shown to be sensitive to immediate linguistic context, a phenomenon called *gestural drift* (Sancier and Fowler, 1997), providing evidence for the immediate effect of reexposure for phonetic aspects.

For the domain of syntax, there are, to our knowledge, only two studies on late bilinguals specifically focusing on reexposure. Chamorro et al. (2016) investigated antecedent preferences for pronominal subjects in Spanish-English bilinguals within the framework of the Interface Hypothesis. They tested two groups of 24 L1 Spanish speakers who had been living in the United Kingdom for a minimum of 5 years. One of these groups had been (re)-exposed to Spanish for at least a week before testing. A control group involved Spanish speakers with very little knowledge of English who had only recently arrived in the United Kingdom (with a mean LoR of 8 weeks). The linguistic material was tested in an offline judgment task and in an online eye-tracking experiment. While there were no differences between the groups in the judgment task, non-exposed attriters showed a lack of online sensitivity for pronoun mismatches in the eye-tracking measures, which distinguished them from both the control group and the recently re-exposed group. Similar results were obtained in a case study on pronominal use in spontaneous speech production by Genevskaja-Hanke (2017), details follow below. The authors of these studies interpret their results as evidence for the conclusion that attrition affects interface structures without causing permanent changes to knowledge representations (in the sense of language competence) in late bilinguals. The attested changes in attrition are temporary instead and we use the terms *temporary* and *temporariness*, when referring to those in the following. However, as most attrition studies, the two studies are cross-sectional and not longitudinal.

In sum, the picture arising from the literature reviewed here, is that bilingual subjects are sensitive to context of use in a much more immediate fashion than previously thought. However, when and what is influenced by the linguistic context is not yet perfectly clear. All we know is that attrition is most likely to arise “... in those instances where the two languages are sufficiently similar to allow some kind of spillover” (Schmid and Köpke, 2017, p. 653). This is specifically the case for domains where the same linguistic features are present in both languages but are subject to distributional variation of some kind. Since the present investigation was aimed at capturing the evolution of linguistic behavior in L1 at different points of time, we focussed on the alternation of overt vs. null



pronominal subjects in speech, a linguistic phenomenon that has previously been shown to be sensitive to variation in different populations (monolinguals, bilinguals, second language users and attriters). Moreover, the languages investigated here, Bulgarian and German, are a promising combination with respect to this phenomenon, as outlined in the next section.

## LINGUISTIC BACKGROUND

### Previous Research on Overt and Null Pronominal Subjects in L1 Attrition

The alternation of null and overt subjects at the syntax-discourse interface has been investigated for different language combinations in recent research on language attrition. Sorace (2005) tested near-native L2 English speakers with L1 Italian on subject use after prolonged exposure to English. These speakers overproduced overt subjects, performing significantly different from Italian monolinguals in topic continuity contexts (see Tsimpli et al., 2004 for details). The same pattern was also found for L2 speakers of Italian (same language combination) and the attested difficulties have been termed *residual optionality* for L2 speakers and *emerging optionality* for speakers with L1 attrition. This led Sorace to the postulation of the *Interface Hypothesis* as a unified framework of bilingualism, treating L2 acquisition, bilingual L1 acquisition and L1 attrition alike (Sorace and Filiaci, 2006 and related work). According to this hypothesis, phenomena that are purely syntactic (at an internal interface) are impervious to attrition and acquirable in L2, while external interface phenomena might lead to persistent deficits in both groups of speakers. In particular, it is the integration of syntactic and discourse properties at the syntax-discourse interface, which is viewed as problematic. The difficulties of the speakers are attributed to either deficient competence or processing but note that representational accounts do not exclude co-occurring processing deficits. In addition, there is a debate on the role of related cross-linguistic differences. Sorace et al. (2009) suggest that this role is minor, because overproduction of the kind in question has not only been attested for speakers of language combinations of a pro-drop and a non-pro-drop language like Italian-English but also for Spanish-Italian bilinguals, who are speakers of two pro-drop languages. However, there has been recent evidence for differences across languages in relation to the scope of overt pronouns (see Filiaci, 2010 for Spanish vs. Italian and Prentza and Tsimpli (2013) for Spanish vs. Greek). Crucially, the possible impact of cross-linguistic differences on bilinguals' performance does not exclude but rather enhances co-occurring processing effects.

Looking at more research on pronominal use in L1 attrition, similarly deviant performance has been attested for other language combinations and various interface phenomena (e.g., Tsimpli, 2007; Perpiñán, 2013; Caloi et al., 2018; Di Dimenico and Baroncini, this issue). For instance, Tsimpli (2007) discusses data from two studies on the interpretation and production of postverbal subjects as well as the alternation of null and overt

subjects by L1 Greek speakers with a near-native competence of English, Swedish and German<sup>3</sup>. The results revealed that the attrited speakers performed significantly different from the monolingual controls without L1 attrition. Perpiñán (2013) investigated the use of postverbal subjects in *wh*-movement constructions in Spanish, testing the performance of L1 Spanish L2 English bilinguals with postpuberty L1 attrition. No effects were found for postverbal subjects in *wh*-matrix questions (considered purely syntactic) but for the same type of subjects in embedded sentences, in which discourse plays a role (focus interpretation in particular), attrition effects were attested. Caloi et al. (2018) and Di Dimenico and Baroncini (this issue) also investigated the use of postverbal subjects in relation to the realization of new information focus in L1 Italian L2 German speakers, attesting residual optionality in the competence of attrited and heritage speakers of this language combination. Two of the rare attrition studies providing some results on reexposure also focus on the use of null subjects (see Chamorro et al., 2016; Genevskaja-Hanke, 2017, mentioned above).

### The Two Languages of Investigation – Bulgarian vs. German

Turning our attention to the overt and null subject alternation, in contrast to non-null subject languages like English and semi-null subject languages like German (both non-pro-drop languages), null subject languages like Italian, Greek, Spanish and Bulgarian (all pro-drop or consistent null subject languages) allow for null referential pronominal subjects (labeled *pro*) in addition to overt referential pronominal subjects in finite clauses, giving rise to a pattern of alternation (Bojadziev et al., 1999; Genevskaja-Hanke, 2019, for Bulgarian; Rizzi, 1986; Jaeggli and Safir, 1989; Roberts and Holmberg, 2010, for the other languages listed above). Note that while the term *pro-drop* refers to a special type of null subjects (*pro*), the term *null subject* is not restricted to a particular type of null subject. Thus, while both German and Bulgarian are null subject languages, only Bulgarian is pro-drop<sup>4</sup>. Examples (1a), (1b), and (1c) illustrate referential subject use in Bulgarian and German. In each case there is reference to one 1PSG subject (as the subject of the main clause) and one 3PSG subject (as the subject of the subordinate clause), both referential and definite. The construction with two overt pronominal subjects given in (1b) represents the only grammatical option in non-pro-drop languages like English. Since spoken German allows for null topic subjects clause-initially, the 1PSG subject is also grammatical, compare (1c), but this is due to null topic licensing by a different grammatical mechanism, termed *topic-drop*. Thus German is also a topic-drop language, but note that topic-drop is restricted to the spoken register (e.g., Hamann, 1996; Haegeman, 2013; Trutkowski, 2016). German null topics are subjects and objects that are only licensed in clause-initial position and further need to be recovered through discourse in the same way Chinese

<sup>3</sup>See Tsimpli et al. (2004) for details.

<sup>4</sup>German is classified as a semi-null subject language, since it requires obligatory null non-argumental expletive subjects in all non-clause-initial positions, independent of register (Biberauer, 2010; Roberts and Holmberg, 2010).



null topics are recovered, see examples (2a)–(2c) from Hamann (1996).

- |      |   |                                |
|------|---|--------------------------------|
| (1a) | <i>pro</i> Знам, че <i>pro</i> идва.<br>Znam, che idva<br>*Weiß, dass kommt.<br>*Know that comes.                                 | Bulgarian<br>German<br>English |
| (1b) | Аз знам, че той идва.<br>Az znam, che toj idva<br>Ich weiß, dass er kommt.<br>I know that he comes.                               | Bulgarian<br>German<br>English |
| (1c) | <i>pro</i> Знам, че той идва.<br>Znam, che toj idva<br>Weiß, dass er kommt.<br>*Know that he comes.<br>'I know that he's coming.' | Bulgarian<br>German<br>English |
| (2a) | <i>pro</i> hab' Hans schon angerufen.<br>*have Hans already called<br>'I have called Hans already.'                               | German<br>English              |
| (2b) | hab' ich gestern <i>pro</i> gekauft.<br>*have I yesterday bought<br>'I have bought it yesterday.'                                 | German<br>English              |
| (2c) | Gestern hab *(ich) das gekauft.<br>Yesterday have *(I) that bought<br>'Yesterday, I have bought that.'                            | German<br>English              |

From a cross-linguistic perspective, both languages, Bulgarian and German allow null referential pronominal subjects. However, in Bulgarian these subjects are licensed through pro-drop and are thus unrestricted in their distribution as to clausal position, while in German they are licensed through topic-drop and only occur clause-initially. Furthermore, German null topics are register-dependent and thus a feature of spoken language. Accordingly, overt referential subjects are generally used to a much higher extent in German than they are in Bulgarian so that a possible influence of German would be an increased use of overt subjects<sup>5</sup>. Furthermore and crucially, the overt referential subjects in German overlap with Bulgarian null subjects in contexts of topic continuity – in other words, while German uses overt subjects, Bulgarian uses null subjects in the very same contexts. This further reflects the difference in the scope of overt pronouns between pro-drop and non-pro-drop languages: in the former type of languages overt pronouns carry both + *topic shift* and – *topic shift* features; in contrast, in the latter type of languages, they only carry a + *topic shift* feature since null pronouns are associated with the – *topic-shift* feature, giving rise to a one-to-one mapping pattern for overt and null pronouns (Tsimpli et al.,

<sup>5</sup>Note that we do not completely exclude the possibility that the presence of null topics in German influences bilinguals' performance so that there might be less overproduction of overt subjects in comparison to the case of a non-null subject language like English. Since null topics overlap with null subjects in pro-drop languages in clause-initial contexts, a "non"-difference in subject use between clause-initial and non-clause-initial contexts would be revealing – in pro-drop languages, null subject rates are higher for subordinate clauses, which in combination with clause-initial topic-drop would give rise to a pattern of a more balanced use of null subjects in clause-initial and non-clause-initial contexts. However, this was not attested in our data. This might be further influenced by the fact that topic-drop is a phenomenon of spoken German. Furthermore, a study on near-native L1 Bulgarian L2 German speakers revealed that topic-drop was not acquired so that these speakers transferred the null subjects of their pro-drop L1 to their L2 (Genevskaja-Hanke, 2019).

2004). In addition, the less restrictive grammar is taken to affect the more restrictive grammar, so that for speakers of a pro-drop L1 with a dominant non-pro-drop L2, neutralization of native distinctions toward the less restrictive L2 option sets in.

The overt vs. null alternation pattern in pro-drop languages also depends on discourse and is thus not exclusively grammatically-driven. Hence, subject use is generally dependent on conditions of the syntax-discourse interface. While overt referential subjects are predominantly used in focal and topic shift contexts, null referential subjects occur in topic continuity contexts, compare the Italian examples (3) and (4) from Roberts and Holmberg (2010):

- |     |   |
|-----|---|
| (3) | Il professore <sub>i</sub> ha parlato dopo<br>the professor has spoken after<br>che lui <sub>*i/j</sub> e arrivato.<br>he is arrived<br>'The professor started speaking after he arrived.'        |
| (4) | Il professore <sub>i</sub> ha parlato dopo<br>the professor has spoken after<br>che <i>pro</i> <sub>*i/j</sub> e arrivato.<br>he is arrived<br>'The professor started speaking after he arrived.' |

This gives rise to specific patterns that are strongly preferred by native speakers (see Sorace, 2005 for Italian). In other words, these patterns are a matter of preference rather than categorical behavior so that sometimes overt subjects surface in topic continuity or non-focal contexts. This is probably due to the following: on the one side, both types of constructions, one with an overt and one with a null subject are generally possible in pro-drop languages (recall examples 1a and 1b); on the other side, the type of pronouns has to be considered in relation to their scope<sup>6</sup>. As above mentioned, there are cross-linguistic differences as to the scope of overt pronouns in pro-drop languages (see Filiaci, 2010 for Italian vs. Spanish and Prentza and Tsimpli (2013) for Greek vs. Spanish and Di Domenico and Baroncini, this volume for Italian vs. Greek), but despite these it generally holds that non-native speakers with a non-pro-drop L1 use overt subjects to a significantly higher extent than native speakers. This has been also attested for some postpuberty L1 attrition speakers but as recent studies on reexposure suggest, their attrition might be temporary. After all, the difference between overt subject use of native speakers in comparison to that of non-native speakers is one of degree and can be, e.g., directly read off the rates of the overt and null subject alternation for the language under consideration. For Bulgarian, a distribution of 27% of overt and 73% of null pronominal subjects in speech has been attested

<sup>6</sup>Cardinaletti (2004) discusses an increase of overt subject use in Italian (see also Frascarelli, 2007 for related evidence from corpora on spoken Italian), suggesting that the paradigm and the status of Italian personal pronouns are undergoing a change from "strong" to possibly "weak" in the sense of Cardinaletti and Starke (1994). For instance, the 3rd person *lui* and *lei* are already considered weak and are slowly taking up the slots of the archaic *egli* and *ella*. Since the null pronoun *pro* is also weak, the overt weak form can freely alternate with *pro*, without violating grammar.

(Genevskaja-Hanke, 2017, 2019; see Lorusso et al., 2005, for similar data on Italian and Di Domenico and Baroncini, this issue, for similar data on Italian and Greek)<sup>7</sup>. This information on subject rates is relevant since we use spontaneous speech production in the present study (see also Di Domenico and Baroncini, this issue, for the implementation of similar data).

## Focus of the Present Study

As discussed in detail above, while many researchers agree that language dominance changes repeatedly over the lifespan (e.g., Silva-Corvalan and Treffers-Daller, 2015), studies generally focus on the first shift of language dominance that may arise after emigration and there are hardly any studies that take into account reexposure to a formerly attrited language. Reexposure is possibly neglected, because of the general assumption that L1 attrition in adults is a slow process, arising after decades of non-use, and also because of the difficulty to conduct longitudinal research. But if we want to make the picture of the processes at play in bilingual development more complete, we need more longitudinal data taking into account reexposure to a formerly attrited language.

The aim of the present study is to help modestly fill this gap by means of a detailed examination of the effects of changes in language environment and reexposure. We assume that this will allow us to contribute to a better understanding of the interplay between dominance and attrition since we adopt a psycholinguistic approach considering both dominance shift and attrition as modifications of the availability of linguistic structures for ongoing language processing. The time scales involved in the changes in availability of lexical items have already been documented to a certain extent, while data on similar processes with respect to grammatical structures are crucially needed. We want to know whether sentence processing strategies may show similar sensitivity to language exposure and use, and explore the possible temporariness of these changes.

The present study provides data from a longitudinal study of a late Bulgarian–German bilingual, investigated at four different points of her bilingual development. The focus is on the use of overt and null pronominal subjects that has proved to be sensitive in the context of language contact and bilingual development, recall the attested overproduction. We assume that one should be able to capture even subtle changes in overt vs. null subject alternation patterns after short periods of reexposure. Since the data used here is spontaneous speech production, this will further allow us to add more ecological data to the mostly experimental data obtained in previous studies.

## MATERIALS AND METHODS

We investigated pronominal use in the spontaneous speech of a bilingual speaker of the language combination L1 Bulgarian L2 German. She is a late bilingual and modifications in her use of pronominal subjects have been attested in a previous study (Genevskaja-Hanke, 2017). That study focused on the

rates of overt and null subjects used by the speaker under consideration of context (topic shift, topic continuity and focal contexts) and aimed at spotting possible overproduction of overt pronominal subjects (see sections Procedure and Review of the Results at Investigation Point 1 for details). Its findings showed that the participant produced significantly more overt pronominal subjects than monolingual Bulgarian speakers in a first investigation, but returned to performance within the monolingual range after 2 weeks of vacation in an L1 environment. In the present study, we aimed at gathering further data on the linguistic trajectory of this bilingual subject and added a second data collection point 5 years later, with another reexposure situation in a follow-up design. The merits of this rather untypical case study are its rare status of being longitudinal, the significant length of L2 exposure in combination with limited L1 contact and the specific language background of a Slavic pro-drop language and a Germanic non-pro-drop but semi-null subject language, additionally allowing for null topics in its spoken register, which has not been studied in the context of L1 attrition so far.

## Participants

Eleven adult Bulgarian native speakers were recorded while conversing, one bilingual speaker and 10 monolingual speakers. The monolingual data is from Genevskaja-Hanke (2019). We used a questionnaire on language background for all participants, which included questions related to age of initial exposure, language proficiency, duration and extent of language influence, languages of family members and friends as well as to previous and current language use, countries of residence (for the lifespan), schooling and age<sup>8</sup>. In relation to language use, detailed information was gathered as to patterns and extent of use at home, at work, with different conversational partners etc.

The bilingual speaker, who is our test participant, grew up as a monolingual speaker of Bulgarian with the exception of learning English in a school setting from grade 5 to grade 7. Both her L2s, German and English, were acquired after puberty and thus fall into the domain of late second languages. She is a proficient speaker of German as a foreign language (attested by a certificate for the foreign and second language proficiency level C1 of the Common European Framework of Reference) and she majored in sociology in Germany. At investigation point one (IP1), she was 32 years old and has lived in the target language country for 12 years. By the time of this first investigation point, she had extremely limited contact to her native language Bulgarian (short stays in Bulgaria roughly every second year and overall rare contact to the language). According to the analysis of the questionnaire data and according to the criteria for near-nativeness as defined by Tsimpli et al. (2004), her competence in German is considered near-native (see White and Genesee, 1996; Tsimpli et al., 2004 for a definition). In other words, she has reached ultimate attainment of her L2 and her German

<sup>7</sup>The rates are from Genevskaja-Hanke (2019), details follow below.

<sup>8</sup>The questionnaire was developed on the basis of the NOWETAS adaptation of the Beirut-Tours Questionnaire on child bilingualism (Paradis J., 2007). It is available in Bulgarian and German, but see the **Supplementary Material** for a list of selected questions in English. For the bilingual, the questionnaire was applied at each investigation point, prior to and after reexposure.

**TABLE 1** | Overview of the recordings of the bilingual.

Investigation point	IP1		IP2	
Recording	1A TC	1B HC	2A TC	2B HC
Country of recording	Germany	Bulgaria	Germany	Bulgaria
Year of recording	2012	2012	2017	2017
Time between recordings	2, 5 weeks		3 weeks	

is hardly distinguishable from that of native speakers without linguistic scrutiny<sup>9</sup>. Following the criteria of the dominance test presented above, including exposure, patterns of use, proficiency and attitudes and on the basis of the answers provided to the questionnaire, German is considered her dominant language. Five years after IP1, there was a second investigation point (IP2) (see **Table 1** for details). Three years before this second investigation point, she married a Bulgarian who moved to Germany and started learning German as a second language himself, which strongly affected her daily language use toward a much more balanced pattern of use for the two languages.

As for the 10 monolingual speakers of Bulgarian (our control group), all participants are considered predominantly monolingual since they had some limited foreign language instruction at school (several decades before recording), which is typical for people born and raised in Europe in their age (mean age 50, age range 30–67)<sup>10</sup>. All 10 are native speakers of Bulgarian, Bulgarian residents born to Bulgarian monolinguals in Eastern Bulgaria (region of Varna), with no or only vacation stays in foreign countries. All participants had either gained a BA degree or completed professional training after graduating from high school.

In relation to data collection, all subjects gave written informed consent in accordance with the declaration of Helsinki. At the time the research with the monolinguals was planned, the University of Oldenburg did not have a protocol for ethical approval/ethics committee for the humanities. The bilingual speaker gave written consent on anonymity and data handling totally conform to the recommendations of the commission for the evaluation of research consequences and ethics of the Carl-von-Ossietzky University of Oldenburg.

## Procedure

We conducted an exploratory longitudinal study focussing on a single case, compared to a control group. Case studies are particularly indicated in research on the dynamics of developmental processes since these allow the researcher to capture a more fine-grained picture of intra-individual variation over time (e.g., Duff, 2014).

We used spontaneous speech data, which resembles language production in real time. The language of this corpus is informal. Participants were recorded, while conversing with one or more speakers in a naturalistic daily life environment. Each recording lasted 60 min on average. The speakers did not receive any

particular instructions prior to the recordings but were informed that the investigation is on the use of Bulgarian in general. In the interviews, they were asked questions thematically linking the conversation to people so that a considerable amount of referential pronouns is used<sup>11</sup>.

The recordings of the monolinguals were produced in 2011 and transcribed, glossed, translated and analyzed thereafter. The ones of the bilingual speaker were produced in 2012 and 2017, see **Table 1** for details.

For each speaker of the control group, 200 utterances on average were analyzed with one exception – for one speaker we collected several recordings with a total of 1000 utterances with the aim of increasing the reliability of the data. Four recordings of the test participant with a total of approximately 550 utterances were analyzed, two at investigation point one, after 12 years of residence in Germany and two after 17 years of residence in Germany. At each investigation point, there was one recording in the country of residence (the target language country, TC, Germany) and a second one after a 2 weeks stay in the home country (HC, Bulgaria)<sup>12</sup>. We analyzed 13 recordings of the controls (nine recordings of nine individual speakers with a length of 200 utterances each and four recordings of one speaker with a total of 1000 utterances).

In the analysis, overt and null subjects were calculated per speaker and per clausal type under consideration of subject and context type. All relevant contexts were considered: focal, topic shift, and topic continuity contexts. Imperatives were excluded, while cases of subject doubling entered the count as two overt pronouns, which minimally raises the respective rates accordingly. For overt subjects, we calculated separate rates for all occurrences of overt subject material (including DPs and pronouns) and for overt pronouns only per participant in order to increase comparability across recordings. Note that overproduction of overt subjects would necessarily affect the null subject rates and subject use would then overall fall short of the monolingual standard.

## Predictions

Starting with the results of the two studies involving reexposure data (Chamorro et al., 2016; Genevskaja-Hanke, 2017), a clear difference between the time before and after reexposure of the non-dominant, attriting language has been reported. Before reexposure, the performance of the attriters was different from that of non-attrited monolinguals (and further comparable to that of second language speakers), as attested in the studies reviewed above (Tsimpli et al., 2004; Sorace, 2005)<sup>13</sup>. After reexposure, the difference between attrited and non-attrited speakers disappeared and the non-dominant language mirrored the so-called “native standard” or “monolingual norm.” Whether this entails another change of dominance remains to be

<sup>9</sup>With the exception of foreign accent, to which native speakers are very sensitive, see e.g., Tsimpli et al. (2004) and White and Genesee (1996).

<sup>10</sup>Instruction was in Russian, except for one speaker who had English instead due to a change in the schooling system.

<sup>11</sup>Possible questions were: “Tell me about your brother...” and “Did you visit anybody lately?”

<sup>12</sup>The results of IP1 have already appeared in Genevskaja-Hanke (2017).

<sup>13</sup>There is also a study on pronominal use in L1 Bulgarian L2 German speakers, which provides evidence that near-native speakers’ competence falls short of the native standard (Genevskaja-Hanke, 2019). This study was not reviewed above since it is an L2 and not an L1 attrition study.

established. In other words, we generally expect attrition effects to be temporary, which entails that the underlying knowledge representations (or language competence in the sense defined above) will not be affected. Thus, on the basis of the results reported in these studies (including these of IP1), we expect attrition effects for the time before reexposure (for the performance of recording 2A TC) for the present study of the second investigation point, IP2. This prediction is also in accordance with assumptions of the Interface Hypothesis on emerging optionality in L1 attrition. However, since this hypothesis does not predict temporariness of the kind reported in Genevskaja-Hanke (2017) and Chamorro et al. (2016), the nature of the optionality for late L1 attriters possibly needs reconsideration. For the time after reexposure, there are two possibilities – monolingual-like performance due to the increase in accessibility of the language or its dominance as in the case of the second recording of IP1 or performance comparable to that of the first investigation time of IP1, possibly due to the longer period of time of exposure to L2. We prefer the former over the latter possibility, because L1 stability is reached by age 12 (as suggested by Schmid, 2014), because monolingual performance has been also attested for the pronominal alternation at the syntax-discourse interface after reexposure) and since there has been no counter-evidence for attrition effects upon L1 reexposure after comparably long periods of L2 exposure so far (IP2 for our participant lies 17 years after immigration).

## RESULTS

### Review of the Results at Investigation Point 1

We start with a review of the results at IP1 (Genevskaja-Hanke, 2017). The monolingual group mean for overt pronominal subjects lies by 27% range 16–36%,  $SD = 0.05794$ , the group is

normally distributed, according to the statistical analysis carried out.

As for the bilingual participant, at the time of the first recording 1A TC, we found overproduction of overt pronominal subjects in the language data of the test participant<sup>14</sup>. Examples (5) and (6) illustrate her use of overt pronominal subjects in topic continuity contexts. Note that the 3PSG subjects in (5) and the 1PSG subjects in (6) all refer to a continuous topic each, established in previous discourse.

The overall rate of overt pronominal subjects of the bilingual reached 41%, exceeding the upper limit of the non-attrited monolinguals' rate range. This rate is significantly different from the rates of the controls, two-tailed probability  $p = 0.043$ , estimated percentage of normal population falling below individual's score = 99.57% (single case statistics significance test on difference between individual's score and control, Crawford and Garthwaite, 2002).

However, the statistical analysis of the two separate rates, the 1A TC and the 1B HC rate of the attrited speaker, revealed that these two rates are significantly different. The 1A TC recording rate indicates that overt pronominal subjects are used in up to 47% of all cases and it is significantly higher than the controls' mean rates, two-tailed probability  $p = 0.009$ , estimated percentage of normal population falling below individual's score = 97.85% (same statistical procedure as above, Crawford and Garthwaite, 2002). For the recording 1B HC, there was no overproduction of overt pronominal subjects. The overt pronominal subjects rate was 34%, which lies within the range of the controls, and thus shows comparable performance, two-tailed probability

<sup>14</sup>Note that the rates of overt subjects in topic shift contexts of the bilingual were comparable across recordings so that the difference between overt subject rates originates solely due to an increased use of overt subjects in topic continuity contexts. This is in line with previous studies, reporting overproduction of overt subjects in precisely these contexts (see section Previous Research on Overt and Null Pronominal Subjects in L1 Attrition for related information and references).

- (5) В смисъл от една страна е добре, тя знае, че е добре обаче тя вече  
 V smisal ot edna strana e dobre, tya znae, che e dobre obache tya veche  
 In sense from one side *pro* is good, she knows, that *pro* is good but she already  
 'On the one hand it is good, she knows that it's good but we've been away for the

последните 10 години нас ни няма и тя има само 2 деца, няма други и...  
 poslednite 10 godini nas ni nyama i tya ima samo 2 detsa, nyama drugi i...  
 last 10 years *pro* us usCL hasn't and she has only 2 kids, *pro* hasn't others and  
 last 10 years and she only has two children, she doesn't have other (children) and...'

- (6) Аз за първи път във Христена бях в тях, тя в къщи не беше,  
 Az za parvi pat vav Hristena byah v tyah, tya v kashchi ne beshe  
 I for first time in Hristena was in theirs, she in home not was,  
 'This was the first time I ever visited Hristena, she hasn't been to my place so far,

... аз съм била в тях само един път. Се запознах с Деница и с Пламен  
 ...az sam bila v tyah samo edin pat. Se zapoznah s Denitsa i s Plamen  
 ... I am been in theirs only one time. *pro* REFL acquainted with Deniza and with Plamen  
 ... I've been at hers only once. There I got acquainted with Denitsa and with Plamen.'

Аз въобще не ги познавах.  
 Az vaobshche ne gi poznavah  
 I at all not themCL knew.  
 'I didn't know them at all.'

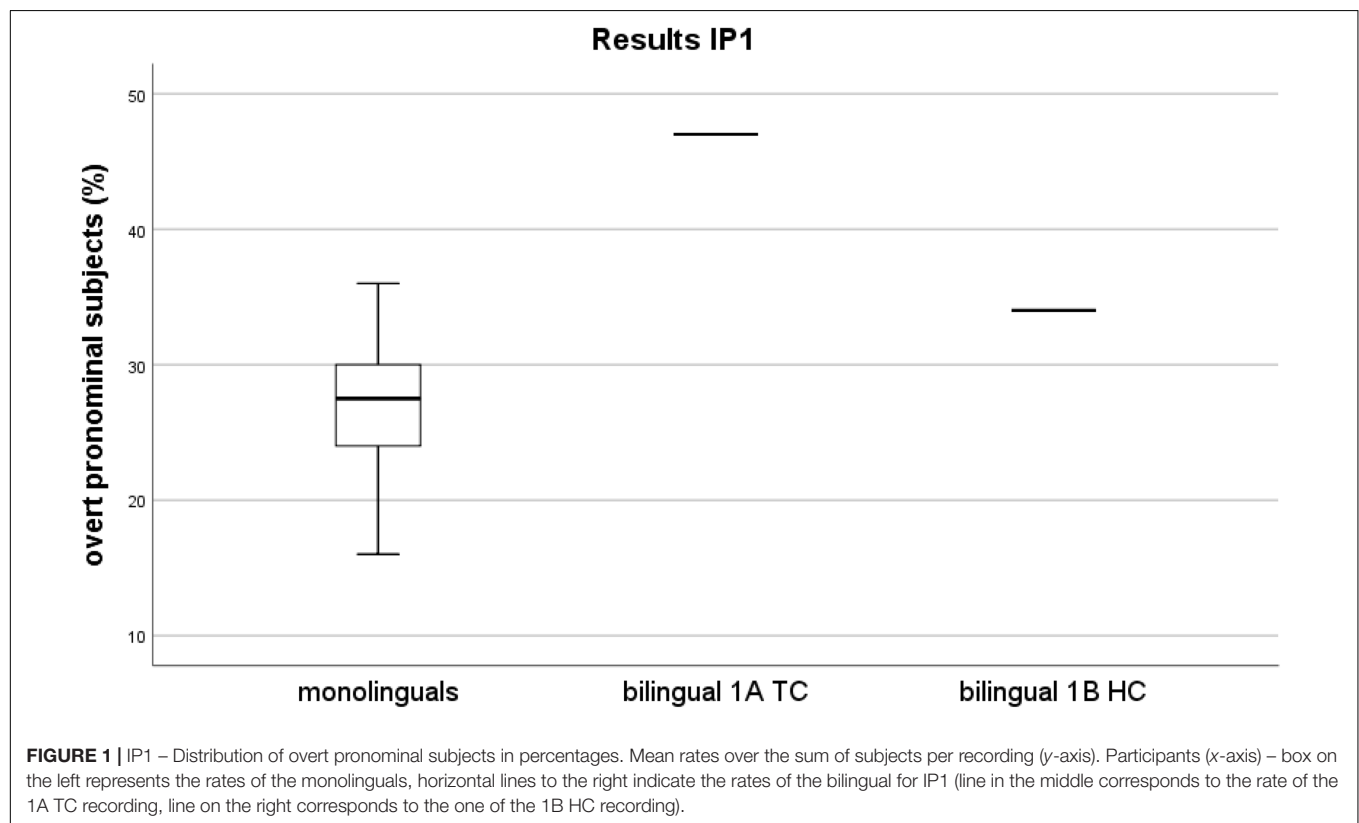


**TABLE 2** | Distribution of overt and null (pronominal) subjects for IP1 vs. IP2.

	Monolinguals (totals)	Bilingual IP1	Bilingual 1A TC	Bilingual 1B HC	Bilingual IP2	Bilingual 2A TC	Bilingual 2B HC
Number of utterances	2909	249	138	111	229	119	110
Number of subjects	3439	291	163	128	266	132	134
Overt subjects	39%	46%	51%	39%	35%	36%	34%
Null subjects	61%	54%	49%	61%	65%	64%	66%
Overt pronominal subjects	27%	41%*	47%*	34%	27%	29%	24%
Null pronominal subjects	73%	59%	53%	66%	73%	71%	76%

\*Significant difference to monolingual group.

Left most column displays the split to number of utterances, subjects, overt and null (pronominal) subjects (all overt vs. overt pronominal only and their null counterparts). Top line indicates the participants' mean rates per recording.



$p = 0.260$ , estimated percentage of normal population falling below individual's score = 86.99% (same type of significance test as above, Crawford and Garthwaite, 2002).

**Figure 1** illustrates the distribution of rates for the L1 group of controls and the mean rates of both recordings of the speaker with L1 attrition at IP1, 1A TC and 1B HC.

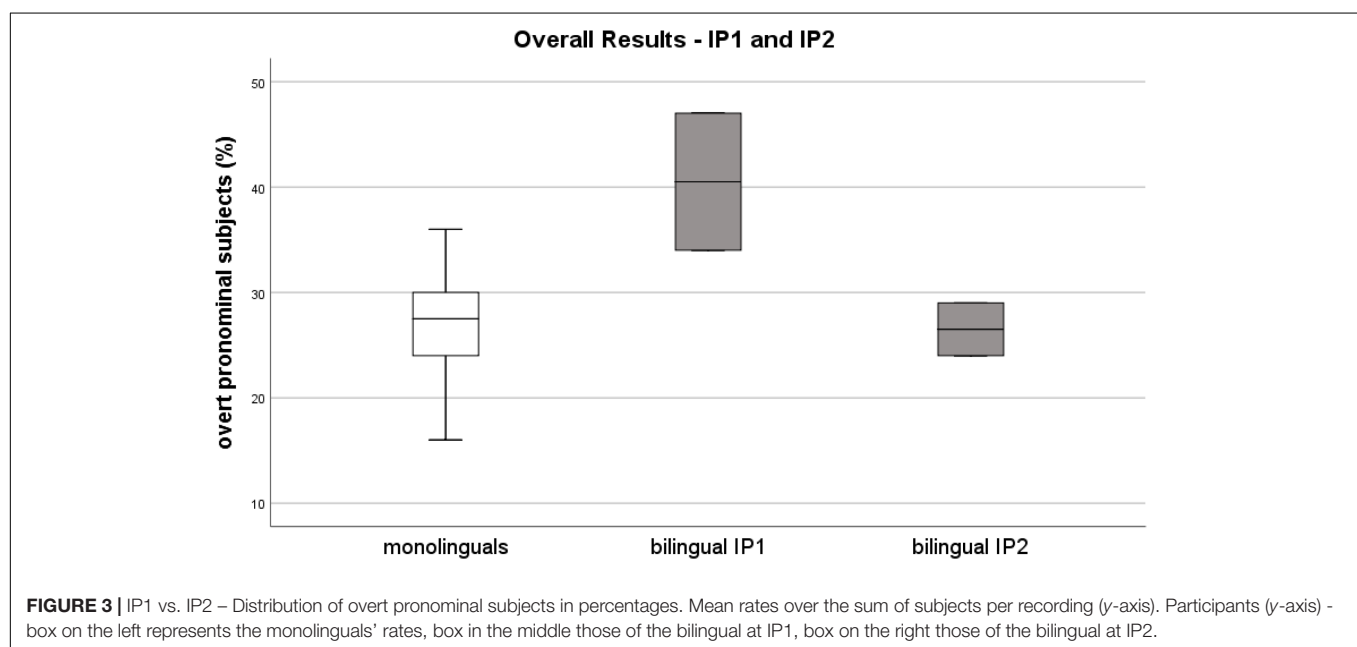
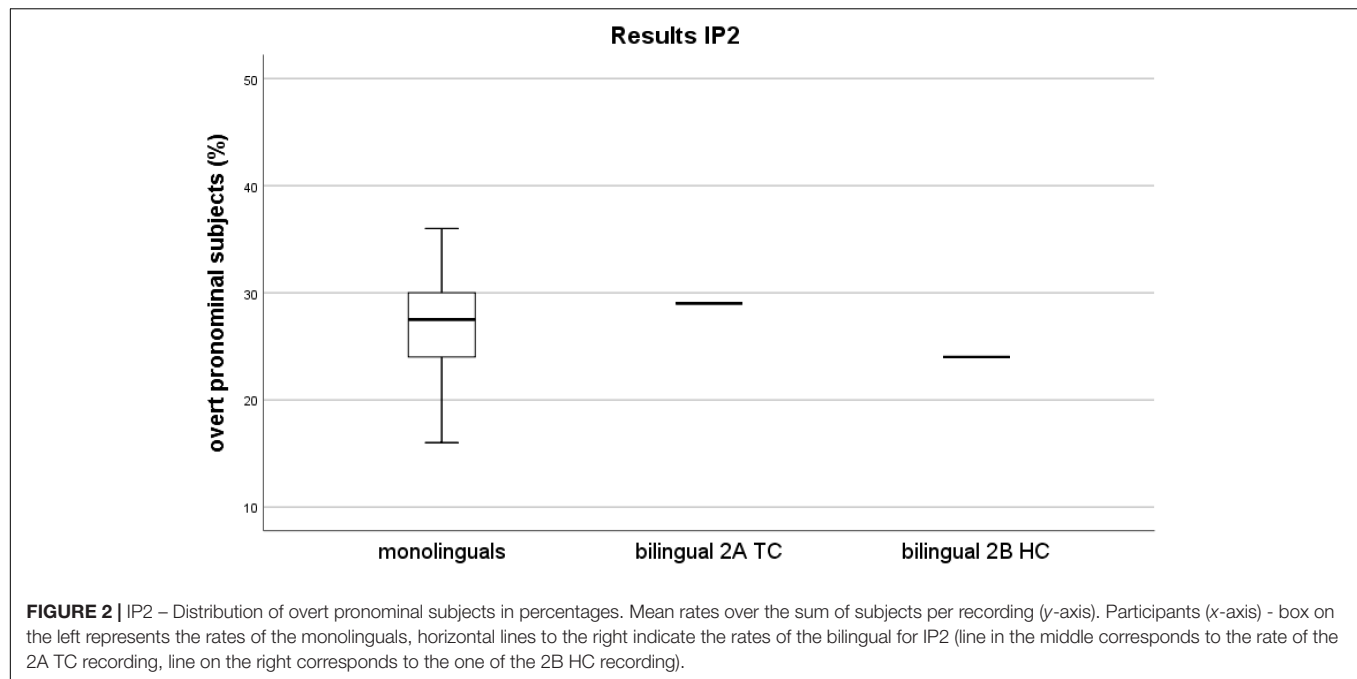
## Results at Investigation Point 2

For investigation point 2, we are using the results of the control group that were already presented in the previous section. The totals of the control group and the 4 rates of the bilingual for both investigation points are displayed in **Table 2**.

At IP2, no overproduction of overt pronominal subjects was attested, neither in the 2A TC recording, nor in the 2B HC recording. Hence the overall rate of both recordings also falls

within the monolinguals' range (two-tailed probability for 2A TC  $p = 0.710$ , estimated percentage of normal population falling below individual's score = 64.29%; for 2B HC,  $p = 0.670$ , estimated percentage of normal population falling below individual's score = 33.39%).

The major difference to the performance at IP1 is the fact that this time both rates of overt pronominal subjects, the 1A TC rate of 29% and the 1B HC rate of 24% fall into the non-attributed monolinguals' rate range. The 1A TC rate is higher than the 1B HC rate, so that this can be interpreted as a similar tendency of a rate drop after reexposure, comparable to that of investigation point one. However, both rates neither differ significantly from one another, nor from those of the monolingual control group (same statistical analyses as those at IP1).



**Figure 2** illustrates the distribution of rates for the L1 group of controls and the mean rates of both recordings of the bilingual speaker at IP2, 2A TC and 2B HC. **Figure 3** illustrates the comparison of the monolingual group and the bilingual speaker at both investigation points.

## DISCUSSION

The overall mean rate for overt pronominal subjects of the L1 Bulgarian L2 German bilingual speaker at the first investigation

point (IP1) revealed significant overproduction of these subjects. This is similar to what has been observed and interpreted as attrition in other studies investigating pronominal use in bilinguals, recall the details given in section Previous Research on Overt and Null Pronominal Subjects in L1 attrition. However, the corresponding overall mean rate 5 years later, at IP2, indicates that the production of overt pronominal subjects differed no longer significantly from the control group data after 17 years of immigration. Additionally, while the overt subject rate of the first recording in the target country (1A TC) was significantly different from the rates of the monolingual controls (which yielded a

further difference for the overall rate of IP1), the overt subject rate of the second recording in the target country (2A TC) was similar to the monolinguals' rates. Thus, contrary to our expectations, no significant differences were measured and accordingly no optionality was attested at this second investigation point after 17 years of immigration, be it in the target or in the home country.

These findings, first of all, point toward the temporariness of attrition phenomena – very similar to what may be observed with respect to dominance shift. The home country recording of each investigation point was made only after few weeks of reexposure to the native language in the home country and the rate of overt pronominal subjects was lower than that of the target country recording each time (albeit non-significantly for the second investigation point), suggesting that a limited amount of extensive exposure to the L1 is sufficient to return to performance within the monolingual speakers' range. While the present data have been obtained with only one bilingual subject, they are perfectly in line with the findings of the cross-sectional study by Chamorro et al. (2016) showing that a group of Spanish immigrants in Great Britain who were immersed during 1 week in an L1 environment performed conform to the native standard, whereas a similar group of immigrants in an L2 environment did not. This means that temporariness in attrition phenomena has now been demonstrated in both longitudinal and cross-sectional data. However, and as above discussed, temporariness in attrition is not predicted by the Interface Hypothesis, which suggests that it still has to be accommodated.

Taken together with evidence from the literature reviewed above, this suggests that peculiarities of performance observed in L1 attrition are probably depending much more on language mode and activation states than on restructuring of linguistic representations (see also Schmid and Köpke, 2017). A processing account for modifications in pronoun use has already been proposed by Gürel (2004). In a study on interpretation of null and overt subject pronouns in embedded clauses in Turkish by Turkish-English bilingual immigrants in North America, she showed that cross-linguistic influence was observed only in those cases, in which Turkish and English allow for similar options in the interpretation of pronouns. Gürel explained this finding with reference to the ATH (Paradis, 1993) predicting that the more frequently used option will be activated more easily when two structures (or lexical items) are in competition, but not when there is no competition because the attriting language has no corresponding structure (as for example in the case of a language with grammatical gender in competition with English, e.g., Bergmann et al., 2015). If the phenomena most commonly observed in attrition studies are due to competition of linguistic options that continue to co-exist in the grammar of the speaker, this clearly means that no structural or representational changes are involved and that processing restrictions may be a promising explanation for the obtained results.

Two factors, however, seem to play a major role in the observed temporariness of preferential processing strategies: immediate language background or language context and age of acquisition. We will discuss these two factors in what follows.

The influence of language context has been demonstrated in the data at two levels. First of all, three of the four recordings

show an effect of the country where the data collection took place. While the first recording in Germany showed significant deviance from the native norms in Bulgarian, data recorded in Bulgaria were within the native range at both investigation points. Such an influence of the immediate language context on performance in a language has been demonstrated repeatedly in recent studies discussed above (e.g., Chang, 2012; Baus et al., 2013; Dussias et al., 2014). It has even been shown that the manipulation of immediate language context has an influence on nonverbal cognitive skills such as control of cognitive interference (Wu and Thierry, 2013). However, the present study also shows that the immediate language environment involves more than just the country of recording: at IP2, performance of the participant in Bulgarian was within the native range also for the recording done in Germany. This can be explained by a change of language use at home. Recall that the participant married a native speaker of Bulgarian 3 years before this investigation point, which made her shift from a quasi exclusive use of German (IP1) to a much more balanced use of both languages (IP2) in her daily life in Germany. The fact that the use of overt and null pronouns in L1 is once again within the native range at this point, after 17 years of immigration, but along with a balanced use of both languages, emphasizes the temporariness of L1 attrition phenomena for late bilinguals. Taken together, the present study contributes to a more general picture suggesting that the language environment must be considered at macro- as well as micro-levels, including, among others, the country of investigation and the specific personal environment at time of investigation at the macro-level (e.g., did the participant receive visits from L1-speakers in the weeks preceding the investigation?) and the languages of the experimenter and the linguistic setting of the task at the micro-level (Wu and Thierry, 2010; Dussias et al., 2014).

However, the language environment is probably not the only factor of influence. What our findings also suggest is that the attrited language of postpuberty L2 speakers may be reactivated relatively fast; within few weeks only. As previously shown, age of acquisition of the L2 is a major factor in determining qualitative and quantitative aspects of attrition (Schmid, 2014). However, for the moment we can only speculate on the role played by age of acquisition with respect to the effects of reexposure. Crucially, studies on reexposure to an attrited language in early bilinguals that could shed further light on the possible temporariness of attrition in younger bilinguals, are not yet available, except for some studies on language relearning in international adoptees (see Oh et al., 2019, for a summary). Nevertheless, there is sufficient evidence for both the observation that dominance changes are frequent and fast in young children (e.g., Slavkov, 2015), and that they are much slower in adults as suggested by the early studies on dominance shift discussed above, reporting periods of L2 immersion of 3–7 years, depending on the tasks used, for adults (Lachman and Mistler-Lachman, 1976; Frenck-Mestre, 1993) and even adolescents (Mägiste, 1979). However, we have to consider two more dimensions. The type of linguistic knowledge involved matters: these studies concern mainly lexical identification and retrieval processes and it is likely that the time scales involved in dominance change will vary for different types of linguistic

knowledge, similar to what has been proposed recently with respect to “critical periods” or other age effects (Birdsong, 2018). But if we assume – as do most studies on language attrition – that the lexicon is most vulnerable to attrition, attaining dominance in L2 for grammatical processing should take even longer than the 3–7 years period observed for lexical processing. Moreover, we have to take into account that these data concern a shift from L1 to L2 dominance, and show that gaining dominance in an L2 over a firmly entrenched L1 takes several years. The present study involves reexposure to the L1 in a late bilingual, and shows that few weeks of immersion are sufficient – if not to reverse dominance – then to at least establish balanced bilingual performance with respect to the grammatical feature investigated here. Hence, for late bilinguals, language status (L1 or L2) is most likely to be a major factor in determining processing ease and permeability to cross-linguistic influence: what seems to remain problematic in L2 acquisition for years (overproduction of overt subjects has been repeatedly attested for L2, as discussed above), may be reestablished within very short time scales after reexposure to a strongly entrenched L1. This points to the fact that different developmental processes as L2 acquisition and L1 attrition need to be considered as distinct, contrary of what is predicted, for instance, by the Interface Hypothesis.

Now, what about our initial question concerning the relationship between dominance shift and attrition? Even though the present study did not focus on dominance as such (only the L1 was investigated), the findings presented here stress the temporary nature of cross-linguistic influence as observed in attrition, affecting language processing and depending on a complex interaction of language exposure and use on the one side and language status as determined by age and order of acquisition on the other side. Furthermore, there is ample evidence that L1 attrition in late bilinguals is generally a processing issue (e.g., discussions in Köpke and Schmid, 2011; Schmid and Köpke, 2017). This provides empirical underpinnings to the idea that attrition and dominance shift are very similar, if not identical processes, involving quantitative but not qualitative differences, which strengthens the idea that we may talk about attrition when the availability of the non-dominant language decreases so much that fluent language processing is becoming more and more difficult (Köpke, 2018). This is obviously not necessarily the case in a bilingual who shows increased use of overt pronominal subjects, as in the present study, unless the person additionally experiences disfluencies in language processing or feels insecure about her language use. The reliance on a processing strategy frequently used in the more dominant language may even be a way to avoid disfluencies in the non-dominant language, a strategy also used by L2 learners (recall that overt pronouns in non-pro-drop languages correspond to both overt and null subjects in pro-drop languages in terms of formal overlap). Further empirical studies of dominance shift at different time scales and in different types of bilinguals are called for in order to challenge these hypotheses, and specifically to investigate whether dominance shift and attrition may be considered as different steps of the same process. An ultimate answer to such a complex question will of course need to be discussed very largely within the field.

## CONCLUSION

The findings of the present study demonstrate the temporariness of attrition phenomena in the domain of pronominal subject use at the syntax-discourse interface. This can be interpreted as evidence for the overall stability of a fully-developed L1 in a late bilingual, as previously proposed for L1 attrition (e.g., Schmid, 2014). Furthermore, L1 attrition in late bilinguals is most likely to arise due to competition of related processing strategies, similar to what definitions of language dominance that go beyond the relative proficiency in each language suggest (Gertken et al., 2014). Viewed in this way, attrition effects appear to be very sensitive to immediate language context at both the macro- and the micro-level. The time scales involved are further dependent on the degree of entrenchment of the language, influenced by age of acquisition of the L2 and the status of the language under investigation (first or second language) to a notable extent.

Thirty years of attrition research have demonstrated very clearly that language systems are dynamic and sensitive to language context. While cross-linguistic influence remains restricted to specific linguistic domains and never entails high error rates, it seems to arise very early in the language contact process, at least as far as specific linguistic structures are concerned. This needs to be investigated and documented carefully in future research employing processing measures under consideration of different types of time scales as well as various settings of immersion and reexposure. Research on language dominance seems a promising way to do this.

## AUTHOR CONTRIBUTIONS

BK developed the main ideas of the manuscript and wrote the sections Introduction, Language Dominance in Research With Bilingual Speakers, Time Scales of Dominance Shift and Attrition, Discussion, and Conclusion. DG-H designed and carried out the study, analyzed its results and wrote the sections Linguistic Background, Materials and Methods, and Results with the exception of the section Focus of the Present Study and parts of the section Procedure. Both authors wrote the abstract, contributed in data interpretation and critically read the manuscript.

## ACKNOWLEDGMENTS

We would like to thank the participants who took part in the study. We are also grateful to Istvan Fekete for his assistance with the statistical analyses and to Cornelia Hamann for her helpful comments on German null subjects.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01963/full#supplementary-material>



## REFERENCES

- Baus, C., Costa, A., and Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychol.* 142, 402–409. doi: 10.1016/j.actpsy.2013.01.010
- Bergmann, C., Meulman, N., Stowe, L. A., Sprenger, S. A., and Schmid, M. S. (2015). Prolonged L2 immersion engenders little change in morphosyntactic processing of bilingual natives. *Neuroreport* 26, 1065–1070. doi: 10.1097/WNR.0000000000000469
- Biberauer, T. (2010). “Semi null-subject languages, expletives and expletive pro reconsidered,” in *Parametric Variation: Null Subjects in Minimalist Theory*, eds T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan (Cambridge: CUP), 153–199.
- Birdsong, D. (2014). Dominance and age in bilingualism. *Appl. Linguist.* 35, 374–392. doi: 10.1093/applin/amu031
- Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Front. Psychol.* 9:81. doi: 10.3389/fpsyg.2018.00081
- Bojadziev, T., Kuzarov, I., and Penchev, J. (1999). *Съвременен български език. Modern Bulgarian Language*. Sofia: Petar Beron.
- Caloi, I., Belletti, A., and Poletto, C. (2018). Multilingual competence influences answering strategies in Italian-German bilinguals. *Front. Psychol.* 9:1971. doi: 10.3389/fpsyg.2018.01971
- Cardinaletti, A. (2004). “Towards a cartography of subject positions,” in *The Structure of CP and IP*, ed. L. Rizzi (New York, NY: OUP), 115–165.
- Cardinaletti, A., and Starke, M. (1994). The typology of structural deficiency: on the three grammatical classes. University of Venice. *Work. Pap. Linguist.* 4, 41–109.
- Chamorro, G., Sorace, A., and Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism* 19, 520–532. doi: 10.1017/S1366728915000152
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *J. Phon.* 40, 249–268. doi: 10.1016/j.wocn.2011.10.007
- Cherciov, M. (2013). Investigating the impact of attitude on first language attrition and second language acquisition from a dynamic systems theory perspective. *Int. J. Biling.* 17, 716–733.
- Clyne, M. (1981). Second generation foreigner talk in Australia. *Int. J. Soc. Lang.* 28, 69–80. doi: 10.1515/ijsl.1981.28.69
- Crawford, J. R., and Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40, 1196–1208. doi: 10.1016/S0028-3932(01)00224-X
- Daller, H. (2011). The measurement of bilingual proficiency: introduction. *Int. J. Biling.* 15, 123–127. doi: 10.1177/1367006910380036
- Daller, M. H., Yildiz, C., de Jong, N. H., Kan, S., and Basbagi, R. (2011). Language dominance in Turkish-German bilinguals: methodological aspects of measurements in structurally different languages. *Int. J. Biling.* 15, 215–236. doi: 10.1177/1367006910381197
- de 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, 331–358. doi: 10.1075/lab.15019.alm
- De Bot, K., and Clyne, M. (1989). Language reversion revisited. *Stud. Second Lang. Acquis.* 11, 167–177. doi: 10.1017/S0272263100000590
- De Bot, K., and Clyne, M. (1994). A 16-year longitudinal study of language attrition in dutch immigrants in Australia. *J. Multiling. Multicult. Dev.* 15, 17–28. doi: 10.1080/01434632.1994.9994554
- De Bot, K., and Schrauf, R. (2009). *Language Development over the Lifespan*. New York, NY: Routledge.
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Clevedon: Multilingual Matters.
- Duff, P. (2014). Case study research on language learning and use. *Annu. Rev. Appl. Linguist.* 34, 233–255. doi: 10.1017/S0267190514000051
- Dunn, A., and Fox Tree, J. (2009). A quick, gradient bilingual dominance scale. *Bilingualism* 12, 273–289. doi: 10.1017/S1366728909990113
- Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *Int. J. Biling.* 3, 355–371. doi: 10.1177/13670069040080031001
- Dussias, P. E., Perrotti, I., Brown, M., and Morales, L. (2014). “Re-learning to parse a first language: the role of experience in sentence comprehension,” in *Proceedings of the 27th CUNY Conference on Human Sentence Processing*, Columbus, OH.
- Dussias, P. E., and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism* 10, 101–116. doi: 10.1017/S1366728906002847
- Ecke, P., and Hall, C. J. (2013). Tracking tip-of-the-tongue states in a multilingual speaker: Evidence of attrition or instability in lexical systems? *Int. J. Biling.* 17, 734–751. doi: 10.1177/1367006912454623
- Filiaci, F. (2010). “Null and overt subject biases in Spanish and Italian: a cross-linguistic comparison,” in *Proceedings of the 12th Hispanic Linguistics Symposium*, eds C. Borgonovo, M. Español-Echevarría, and P. Prévost (Somerville, MA: Cascadia), 171–182.
- Flecken, M. (2011). Assessing bilingual attainment: macrostructural planning in narratives. *Int. J. Biling.* 15, 164–186. doi: 10.1177/1367006910381187
- Flege, J. E., Mackay, I. R. A., and Piske, T. (2002). Assessing bilingual dominance. *Appl. Psycholinguist.* 23, 567–598. doi: 10.1017/S0142716402004046
- Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro. An interface approach to the linking of (null) pronouns. *Nat. Lang. Linguist. Theory* 25, 691–734. doi: 10.1007/s11049-007-9025-x
- French-Mestre, C. (1993). Use of orthographic redundancies and word identification speed in bilinguals. *J. Psycholinguist. Res.* 22, 397–410. doi: 10.1007/BF01074343
- Genevskaja-Hanke, D. (2019). Subject Realization in Bulgarian: Overt and Null Subjects in Bulgarian-German Interlanguage. Doctoral thesis, University of Oldenburg, Oldenburg.
- Genevskaja-Hanke, D. (2017). “Intrapersonal variation in late L1 attrition and its implications for the competence/performance debate,” in *Linguistik im Nordwesten: Beiträge zum 8. Nordwestdeutschen Linguistischen Kolloquium*, eds N. Levkovych and A. Urdze (Bochum: Brockmeyer).
- Gertken, L. M., Amengual, M., and Birdsong, D. (2014). “Assessing language dominance with the bilingual language profile,” in *Measuring L2 Proficiency: Perspectives from SLA*, eds P. Leclercq, A. Edmonds, and H. Hilton (Bristol: Multilingual Matters), 208–225.
- Grosjean, F. (1998). Transfer and language mode. *Bilingualism* 1, 175–176. doi: 10.1017/S1366728998000285
- Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *J. Neurolinguist.* 17, 53–78. doi: 10.1016/S0911-6044(03)00054-X
- Gürel, A. (2017). Is every bilingual an attriter?: The unbearable complexity of defining L1 attrition. *Linguist. Approaches Biling.* 7, 696–699. doi: 10.1075/lab.00007.gur
- Haegeman, L. (2013). The syntax of registers: diary subject omission and the privilege of the root. *Lingua* 130, 88–110. doi: 10.1016/j.lingua.2013.01.005
- Hamann, C. (1996). Null arguments in German child language. *Lang. Acquis.* 5, 155–208. doi: 10.1207/s15327817la0503\_1
- 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
- Hansen, L. (1999). *Second Language Attrition in Japanese Contexts*. Oxford: Oxford University Press.
- Hutz, M. (2004). “Is there a natural process of decay? A longitudinal study of language attrition,” in *First Language Attrition: Interdisciplinary Perspectives On Methodological Issues*, eds M. S. Schmid, B. Köpke, M. Keijzer, and L. Weilemar (Amsterdam: John Benjamins), 189–206.
- Jaeggli, O., and Safir, K. (1989). *The Null Subject Parameter*. Dordrecht: Foris. doi: 10.1007/978-94-009-2540-3
- Jaspaert, K., and Kroon, S. (1992). “From the typewriter of A.L.: a case study in language loss,” in *Maintenance and Loss of Minority Languages*, eds W. Fase, K. Jaspaert, and S. Kroon (Amsterdam: John Benjamins), 137–147. doi: 10.1075/sibil.1.11jas
- Köpke, B. (2018). “First language attrition : from bilingual to monolingual proficiency?,” in *The Cambridge Handbook of Bilingualism*, eds A. De Houwer and L. Ortega (Cambridge: Cambridge University Press).
- Köpke, B., and Schmid, M. S. (2004). “First language attrition: The next phase,” in *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*,

- eds M. S. Schmid, B. Köpke, M. Keijzer, and L. Weilemar (Amsterdam: John Benjamins), 1–43.
- Köpke, B., and Schmid, M. S. (2011). L'attrition de la première langue en tant que phénomène psycholinguistique. *Lang. Interact. Acquis.* 2, 197–220. doi: 10.1075/lia.2.2.02kop
- Lachman, R., and Mistler-Lachman, J. (1976). Dominance lexicale chez les bilingues. *Bull. Psychol.* 15, 281–288.
- Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing access to the native language while immersed in a second language. *Psychol. Sci.* 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x
- Lorusso, P., Caprin, C., and Guasti, M. (2005). "Overt subject distribution in early Italian children," in *A Supplement to the Proceedings of BUCLD*, eds A. Brugos, M. Clark-Cotton, and S. Ha (Somerville MA: Cascadilla Press).
- Mägiste, E. (1979). The competing language systems of the multilingual: a developmental study of decoding and encoding processes. *J. Verbal Learn. Verbal Behav.* 18, 79–89. doi: 10.1016/S0022-5371(79)90584-X
- Mayr, R., Jones, D., and Mennen, I. (2014). "Speech Learning in bilinguals: consonant cluster acquisition," in *Advances in the Study of Bilingualism*, eds E. M. Thomas and I. Mennen (Bristol: Multilingual Matters), 3–24.
- Mehotcheva, T., and Köpke, B. (2019). "L2 attrition," in *The OUP Handbook of Language Attrition*, eds M. S. Schmid and B. Köpke (Oxford: OUP).
- Montrul, S. (2015). "Dominance and proficiency in early and late bilingualism," in *Language Dominance in Bilinguals*, eds C. Silva-Corvalan and J. Treffers-Daller (Cambridge: CUP), 15–35.
- Oh, J., Au, T.-K., Jun, S.-A., and Lee, R. (2019). "Childhood language memory in adult heritage language (re)learners," in *The Oxford Handbook on Language Attrition*, eds M. S. Schmid and B. Köpke (Oxford: Oxford University Press).
- Opitz, C. (2013). A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *Int. J. Biling.* 17, 701–715. doi: 10.1177/1367006912454621
- Paradis, M. (1993). Linguistic, psycholinguistic, and neurolinguistic aspects of "interference" in bilingual speakers: the activation threshold hypothesis. *Int. J. Psycholinguist.* 9, 133–145.
- Paradis, J. (2007). *The Alberta Language Environment Questionnaire (ALEQ)*. Available at: <https://www.ualberta.ca/linguistics/chescentre/questionnaires>
- Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins. doi: 10.1075/sibil.18
- Paradis, M. (2007). "L1 attrition features predicted by a neurolinguistic theory of bilingualism," in *Language Attrition. Theoretical Perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert (Amsterdam: John Benjamins), 121–134.
- Perpiñán, S. (2013). "Optionality in bilingual native grammars," in *First Language Attrition*, eds M. S. Schmid and B. Köpke (Amsterdam: John Benjamins), 127–156.
- Pierce, L., Genesee, F., and Klein, D. (2019). "Language loss and language learning in internationally adopted children: evidence from behaviour and the brain," in *The Oxford Handbook on Language Attrition*, eds M. S. Schmid and B. Köpke (Oxford: Oxford University Press).
- Prentza, A., and Tsimpli, I. M. (2013). On the optionality in L2 pronominal production and interpretation. What (more) can VP-coordination structures tell us? *Eurosla* 13, 22–46. doi: 10.1075/eurosla.13.04pre
- Rizzi, L. (1986). Null objects in Italian and the theory of pro. *Linguist. Inq.* 17, 501–557.
- Roberts, I., and Holmberg, A. (2010). "Parameters in minimalist theory," in *Parametric Variation: Null Subjects in Minimalist Theory*, eds T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan (Cambridge: CUP), 1–57.
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *Int. J. Biling.* 13, 145–155. doi: 10.1177/1367006909339814
- Sancier, M. L., and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *J. Phon.* 25, 421–436. doi: 10.1006/jpho.1997.0051
- Schmid, M. S. (2007). "The role of L1 use for L1 attrition," in *Language Attrition. Theoretical perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert (Amsterdam: John Benjamins), 135–154. doi: 10.1075/sibil.33.10sch
- Schmid, M. S. (2014). The debate on maturational constraints in bilingual development: a perspective from first language attrition. *Lang. Acquis.* 21, 386–410. doi: 10.1080/10489223.2014.892947
- Schmid, M. S., and Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguist. Approaches Biling.* 7, 637–667. doi: 10.1075/lab.17058.sch
- Schmid, M. S., and Yilmaz, G. (2018). Predictors of language dominance: an integrated analysis of first language attrition and second language acquisition in late bilinguals. *Front. Psychol.* 9:1306. doi: 10.3389/fpsyg.2018.01306
- Seliger, H. W., and Vago, R. M. (1991). "The study of first language attrition: an overview," in *First Language Attrition*, eds H. W. Seliger and R. M. Vago (Cambridge: Cambridge University Press), 3–16. doi: 10.1017/CBO9780511620720.001
- Sharwood Smith, M. (1983). "On explaining language loss," in *Language Development at the Crossroads*, eds R. Felix and H. Wode (Tübingen: Gunter Narr Verlag), 49–59.
- Silva-Corvalan, C., and Treffers-Daller, J. (2015). *Language Dominance in Bilinguals. Issues of Measurement and Operationalization*. Cambridge, MA: Cambridge University Press.
- Slavkov, N. (2015). Language attrition and reactivation in the context of bilingual first language acquisition. *Int. J. Biling. Educ. Biling.* 18, 715–734. doi: 10.1080/13670050.2014.941785
- Sorace, A. (2005). "Selective optionality in language development," in *Syntax and Variation*, eds L. Cornips and K. Corrigan (Amsterdam: Benjamins), 55–80. doi: 10.1075/cilt.265.04sor
- 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
- Sorace, A., Serratrice, L., Filiaci, F., and Baldo, M. (2009). Discourse conditions on subject pronoun realization: testing the linguistic intuitions of older bilingual children. *Lingua* 119, 460–477. doi: 10.1016/j.lingua.2008.09.008
- Stolberg, D., and Münch, A. (2010). "Die Muttersprache vergisst man nicht" – or do you? A case study in L1 attrition and its (partial) reversal. *Bilingualism* 13, 19–31. doi: 10.1017/S1366728909990332
- Tsimpli, I. M. (2007). "First language attrition from a minimalist perspective: interface vulnerability and processing effects," in *Language Attrition. Theoretical Perspectives*, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. Dostert (Amsterdam: John Benjamins), 83–98.
- Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *Int. J. Biling.* 15, 147–163. doi: 10.1177/1367006910381186
- Trutkowski, E. (2016). *Topic Drop and Null Subjects in German. Linguistics and Philosophy*. Berlin: De Gruyter. doi: 10.1515/9783110446173
- Tsimpli, I. M. (2017). Crosslinguistic influence is not necessarily attrition. *Ling. Approaches Biling.* 7, 759–762. doi: 10.1075/lab.00021.tsi
- Tsimpli, I. M., Sorace, A., Heycock, C., and Filiaci, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *Int. J. Biling.* 8, 257–277. doi: 10.1177/13670069040080030601
- Unsworth, S. (2015). "Amount of exposure as a proxy for dominance in bilingual language acquisition," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalan and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173.
- Wei, L. (ed.). (2007). "Dimensions of bilingualism," in *The Bilingualism Reader*. (London: Routledge), 3–24.
- White, L., and Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Lang. Res.* 12, 233–265. doi: 10.1177/026765839601200301
- Wu, Y. J., and Thierry, G. (2010). Investigating bilingual processing: the neglected role of language processing context. *Front. Psychol.* 1:178. doi: 10.3389/fpsyg.2010.00178
- Wu, Y. J., and Thierry, G. (2013). Fast modulation of executive function by language context in bilinguals. *J. Neurosci.* 33, 13533–13537. doi: 10.1523/JNEUROSCI.4760-12.2013

**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 Köpke and Genevskaja-Hanke. 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.



# Morpho-Syntactic Abilities of Unbalanced Bilingual Children: A Closer Look at the Weaker Language

Natalia Meir<sup>1,2\*</sup>

<sup>1</sup> Department of English Literature and Linguistics, Bar Ilan University, Ramat Gan, Israel, <sup>2</sup> Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel

## OPEN ACCESS

### Edited by:

Dobrinka Genevska-Hanke,  
University of Oldenburg, Germany

### Reviewed by:

Lina Abed Ibrahim,  
University of Oldenburg, Germany  
Solveig Chilla,  
University of Flensburg, Germany

### \*Correspondence:

Natalia Meir  
natalia.meir@gmail.com

### Specialty section:

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

**Received:** 27 March 2018

**Accepted:** 09 July 2018

**Published:** 13 August 2018

### Citation:

Meir N (2018) Morpho-Syntactic Abilities of Unbalanced Bilingual Children: A Closer Look at the Weaker Language. *Front. Psychol.* 9:1318. doi: 10.3389/fpsyg.2018.01318

Previous studies evaluating morpho-syntactic abilities in the Weaker Language of unbalanced bilingual children are scarce; and they bring inconclusive evidence on the nature of the Weaker Language development. The current study looked into morpho-syntactic profiles of bilingual Russian–Hebrew speaking children in the Weaker Language [the Weaker Heritage Language (HL-Russian) and the Weaker Societal Language (SL-Hebrew)] as compared to balanced bilinguals, unbalanced bilinguals in the Dominant Language and bilinguals with Specific Language Impairment (SLI). Four groups of bilingual children aged 5;5–6;5 participated: unbalanced bilinguals with the Weaker HL-Russian and the Dominant SL-Hebrew (HL-weak:  $n = 39$ ), unbalanced bilinguals with the Weaker SL-Hebrew and the Dominant HL-Russian (SL-weak:  $n = 19$ ); balanced bilinguals (BB:  $n = 38$ ), and bilinguals with SLI (biSLI:  $n = 23$ ). Children's morpho-syntactic abilities in both languages were investigated using LITMUS (Language Impairment Testing in Multilingual Settings) Sentence Repetition Tasks (based on Marinis and Armon-Lotem, 2015). Quantitative analysis of morpho-syntactic abilities showed that unbalanced bilinguals scored lower in the Weaker Language as compared to balanced bilinguals and unbalanced bilinguals in the Dominant Language, yet, higher than bilinguals with SLI. Error patterns were similar across bilingual groups with TLD and could be traced to cross-linguistic influence. By contrast, error profiles of unbalanced bilinguals in the Weaker Language and bilinguals with SLI bore fundamental differences. Whereas unbalanced bilinguals in the Weaker Language opted for complex structures, relying on the available resources from the Dominant Language; bilinguals with SLI simplified complex syntactic structures. To conclude, the study shows that the Weaker Language of unbalanced bilinguals with TLD develop qualitatively similarly to the languages of balanced bilinguals and the Dominant Language in unbalanced bilinguals, albeit delayed or influenced by the Dominant Language to a larger extent. Conversely, the study brings evidence that linguistic profiles of unbalanced bilinguals with TLD in the Weaker Language and bilinguals with SLI differ, pointing at a deviant pattern of acquisition in children with SLI.

**Keywords:** morpho-syntax, unbalanced language development, the Weaker Language, delay, deviance, Specific Language Impairment (SLI)



## INTRODUCTION

Linguistic abilities of bilingual children with typical language development (TLD) are unevenly distributed within and across the two languages (Kohnert, 2010). Typically, most bilingual children have one language which is more dominant (i.e., stronger, more preferred) than the other language. This has been noted for simultaneous bilingual children (those bilinguals who are exposed to both of their languages early in childhood) and sequential bilingual children [those who first acquire the Heritage Language (HL) and then are exposed to the Societal Language (SL)] (e.g., Pearson et al., 1993; Schlyter, 1994; Müller and Kupisch, 2003; Bernardini and Schlyter, 2004). Language dominance as well as language preference changes over the life span of bilinguals (e.g., De Houwer, 1990; Montrul, 2008; Gathercole and Thomas, 2009; Polinsky, 2018). For example, in sequential bilingual acquisition, the SL usually starts as the Weaker Language and often becomes the Dominant Language over time. Conversely, the HL starts as the Dominant language and often becomes the Weaker Language as the SL gains dominance. Previous studies report contradicting findings on the nature of the Weaker Language development. Some studies suggest that unbalanced bilinguals in the Weaker Language show similar trajectories to the ones observed in monolingual children, balanced bilinguals (BB) and unbalanced bilinguals in the Dominant Language, yet this pattern is delayed (e.g., Müller and Kupisch, 2003; Bernardini and Schlyter, 2004; Antonova Ünlü and Li, 2016, 2017, 2018). Conversely, some studies show that the Weaker Language development does not follow the monolingual trajectory, i.e., it resembles adult L2 acquisition or it is influenced by the Dominant Language. Numerous studies have shown that morpho-syntactic abilities of bilingual children are susceptible to cross-linguistic influence and bilinguals diverge from monolingual baseline grammars (Müller and Hulk, 2001; Paradis and Navarro, 2003; Argyri and Sorace, 2007; Kupisch, 2007; Meir et al., 2017; Sorace and Serratrice, 2009). The current study does not aim to compare bilingual children to monolingual “golden” standards, rather it is devised to investigate different types of bilingual language development: balanced versus unbalanced, typical versus atypical. These patterns of bilingual language development are investigated in Russian–Hebrew speaking bilingual children. Russian–Hebrew bilingualism offers a unique opportunity to test cross-linguistic influence since some morpho-syntactic properties are configured similarly in both languages (e.g., verbal inflections), while some properties vary across the two languages (e.g., case morphology, aspectual marking, definiteness).

The goal of the current study is twofold. First, it aims to fill the gap created by “the weak interest in the Weaker Language” (Bernardini, 2017). The study investigates morpho-syntactic skills of two groups of unbalanced bilinguals: bilinguals with the Weaker Heritage Language (HL-Russian) and the Weaker Societal Language (SL-Hebrew). Second, the study aims to add to the delay-versus-deviance debate by comparing linguistic profiles of unbalanced bilingual children with TLD and bilingual children with Specific Language Impairment (SLI). This comparison

is intended to unravel the underlying nature of grammatical representations in the two populations.

To evaluate morpho-syntactic abilities, children’s performance on LITMUS (Language Impairment Testing in Multilingual Settings) Sentence Repetition Tasks (based on Marinis and Armon-Lotem, 2015) were administered in both languages of bilingual children (HL-Russian and SL-Hebrew). Sentence Repetition tasks are widely used to assess morpho-syntactic abilities of monolingual and bilingual children. Sentence Repetition tasks have been shown to be highly effective in discriminating children with typical and atypical language development in monolingual and bilingual populations (e.g., Conti-Ramsden et al., 2001; Archibald and Joanisse, 2009; Klem et al., 2015; Meir et al., 2016; Antonijevic et al., 2017; Gavarró, 2017; Hamann and Abed Ibrahim, 2017; Theodorou et al., 2017; Fleckstein et al., 2018; among many others). In the following subsections, quantitative and qualitative characteristics of the Weaker Language of unbalanced bilinguals are discussed as compared to BB and unbalanced bilinguals in the Dominant Language. Second, the trajectory of the Weaker Language development in unbalanced bilinguals with TLD is discussed in terms of delay and deviance. Third, studies on atypical language development in children with SLI are reviewed. Finally, specific research questions and predictions for the current study are presented.

## The Weaker Language vs. the Dominant Language: Quantitative Characteristics

Determining the Weaker and the Dominant Language in bilinguals is not an easy task and it poses great challenges to linguists, educators and speech pathologists. Meisel (2007) defines a language as “weak” or “non-dominant” based on input and output characteristics, on the one hand, and language skills, on the other hand. It is suggested that the Weaker Language is (a) rarely actively used, (b) the other language is strongly preferred over an extended period of time, and (c) the development of the Weaker Language is less advanced than that of the other language(s).

Previous studies rely on quantitative differences between the two languages of a bilingual child. Quantitative discrepancies in scores across the two languages are viewed as a token of unbalanced bilingual language development. For example, many studies, especially those on younger bilinguals, use mean length of utterance (MLU) and directionality of code-mixing as indices of language dominance (e.g., Schlyter, 1994; Jisa, 2000; Bernardini and Schlyter, 2004). Some studies determine language dominance based on direct language proficiency scores (e.g., Paradis et al., 2003; Iluz-Cohen and Armon-Lotem, 2013). Children who obtain higher scores in one language and lower scores in the other language are labeled as unbalanced bilinguals, higher scores in the language signify the Dominant Language, while lower scores are viewed as a sign of the Weaker Language. Other studies rely on quantitative differences in exposure and output characteristics (e.g., Thordardottir, 2011; Hoff et al., 2012; Gathercole et al., 2014; Unsworth, 2016; de Almeida et al., 2017). For example, Hoff et al. (2012) used estimates of exposure



at home to determine dominance of Spanish–English toddlers. Children who had above 70% of exposure to the language at home, were labeled as dominant in that language. In the study of de Almeida et al. (2017) on bilinguals with SL-French, several measures were used to compute a dominance score: exposure in each of the child's languages, Age of Onset of bilingualism (AoO), frequency of early exposure, diversity of early contexts of exposure, Length of Exposure (LoE), present use of each language at home, present use during different activities and with friends, and number of years the child has spent in elementary school. Some studies combine direct and indirect indices of language dominance, i.e., look at the discrepancies in the proficiency scores and discrepancies in parental ratings of children's language proficiency (e.g., Gutiérrez-Clellen et al., 2006; Bedore et al., 2011).

As demonstrated above, different measures are used in the literature to determine language dominance in bilingual children. Different indices might result in different classification labels for bilinguals. This has been demonstrated by Bedore et al. (2012) which compared indices of children's input and output characteristics, based on parental questionnaires, as well as language skills in HL-Spanish and SL-English among 1029 Spanish–English bilinguals with different levels of language dominance: functional monolingual English, bilingual English-Dominant, BB, bilingual Spanish-Dominant and functional monolingual Spanish. The child's current language use was found to be the stronger predictor of the children language performance as measured by direct assessment of language skills in HL and SL. Lust et al. (2016) showed a discrepancy between difference indices of language exposure for Korean–English bilinguals suggesting that parental reports should be supplemented by direct measures of language assessment.

To sum up, quantitative discrepancies across different measures (e.g., MLU, directionality of code-mixing, parental ratings, exposure patterns, language scores (vocabulary and morpho-syntax) are used to determine language dominance in bilinguals. Yet, these quantitative measures do not shed light on qualitative characteristics of the Weaker Language. The next subsection will discuss qualitative properties of the Weaker Language of unbalanced bilinguals and trajectories of the Weaker Language development with the main focus on morpho-syntax.

## Morpho-Syntactic Abilities in the Weaker Language of Bilinguals With TLD: Delayed or Deviant

The Weaker and the Dominant Languages of a bilingual child vary quantitatively as it has been demonstrated in the previous section. Yet, with respect to the qualitative differences there is no agreement. On the one hand, linguistic profiles of unbalanced bilinguals in the Weaker Language resemble those of BB and bilinguals in the Dominant Language. Some studies even show that the Weaker Language development is qualitatively similar to the one of monolinguals. Conversely, some studies show that error patterns in the Weaker Language differ from those of monolinguals, BB and bilinguals in the Dominant Language. This gave rise to two competing hypotheses on the nature of the

Weaker Language development: the Delay Hypothesis and the Deviance Hypothesis.

Similarities in error profiles of unbalanced bilinguals in the Weaker Language and other groups of children (e.g., monolinguals, BB and unbalanced bilinguals in the Dominant Language) provide support for the Delay Hypothesis. For example, Müller and Kupisch (2003) showed that despite quantitative differences in the development of the Weaker Languages of French–German unbalanced bilinguals, the Weaker and the Dominant Languages are qualitatively similar. In the same vein, Bernardini and Schlyter (2004) noted that the developmental trajectory of the Weaker Language of simultaneous Swedish–Italian/German bilingual children followed the same milestones as in the Dominant Language, but the lexical realization was delayed. Several recent studies investigating language development in a simultaneous bilingual child with the Weaker HL-Russian and the Dominant SL-Turkish show that despite reduced input in HL-Russian, the acquisition of grammatical categories in the Weaker HL-Russian (e.g., aspect marking, case morphology and grammatical gender assignment) follows the same pattern as in monolingual acquisition (Antonova Ünlü and Li, 2016, 2017, 2018).

In contrast, there is also evidence that morpho-syntactic abilities of unbalanced bilinguals in their Weaker Language differ not only quantitatively, but also qualitatively as compared to monolingual children (Müller and Hulk, 2001; Paradis and Navarro, 2003; Argyri and Sorace, 2007; Kupisch, 2007; Sorace and Serratrice, 2009; Ringblom, 2012; Meir et al., 2017; Dobrova and Ringblom, 2018). This line of research supports the Deviance Hypothesis, suggesting that the Weaker Language of bilinguals is influenced by the Dominant Language. For example, Ringblom (2012), based on the longitudinal data of a simultaneous bilingual Russian–Swedish child, concluded that the development of the Weaker Language (HL-Russian) did not always follow a monolingual trajectory and was strongly influenced by the Dominant Swedish. The acquisition of rich morphology in the Weaker HL-Russian was reported to be challenging in contact with SL-Swedish which has sparse inflectional morphology. As for complex syntactic development of the Weaker HL-Russian, the errors produced by the child clearly suggest that the Weaker Language heavily relies on the Dominant language. The production of relative clauses in the Weaker HL-Russian was supported by the Dominant SL-Swedish: *eto ja som sdelał eto* 'this I who did this,' where the Russian *wh*-pronoun which should be inflected for case, gender and number is replaced by a Swedish uninflected complementizers *som* (see Dobrova and Ringblom, 2018).

Rodina and Westergaard (2017) showed that Russian–Norwegian bilinguals with two Russian-speaking parents show similar performance to monolinguals on gender agreement/assignment in HL-Russian. However, the bilinguals with the Weaker HL-Russian, who grew in one-parent-one-language families, showed not only a quantitative disadvantage as compared to monolinguals but also a different error profile. Bilinguals with the Weaker HL-Russian predominantly used masculine agreement across the board, this error pattern is neither observed in monolinguals nor in BB.

A recent study by Janssen (2016) investigated the acquisition of nominal morphology by Dutch-Dominant bilinguals with HL-Russian and HL-Polish as their Weaker Languages in comparison with BB and bilinguals in the Dominant Language. Dutch-Dominant bilinguals had more difficulties with case morphology and gender agreement/assignment in their Weaker HL-Polish and HL-Russian as compared to BB and bilinguals in the Dominant Language. Problems with case morphology in HL-Polish and HL-Russian can be attributed to the influence of the Dominant SL-Dutch which does not use nominal inflections. Yet, error profiles were not compared across the bilingual groups.

To sum up, most previous studies addressing the Weaker Language development have compared unbalanced bilinguals to monolinguals (but see Müller and Kupisch, 2003; Bernardini and Schlyter, 2004). Rather than comparing monolingual and bilingual grammars, the current study will probe whether grammatical representations in the Weaker Language are similar/different to those of BB and unbalanced bilinguals in the Dominant Language. Few studies investigated error profiles of unbalanced bilinguals in their Weaker Language as compared to BB and unbalanced bilinguals in the Dominant Language to determine whether the Weaker Language is delayed or deviant from other bilinguals. In this study, deviance is viewed as diverging from other bilingual patterns of acquisition, rather than from monolingual ones. Moreover, the assumption of the current study is that the Weaker Language development in bilinguals with TLD, whether delayed or affected by the Dominant Language, is not disordered. Previous findings show that there are quantitative and qualitative differences between monolingual children with SLI and bilinguals with TLD (e.g., Paradis et al., 2008; Armon-Lotem, 2014). Thus, comparison of linguistic profiles of unbalanced bilinguals in the Weaker Language and bilinguals with SLI will shed light on the developmental trajectories of both populations.

## Morpho-Syntactic Abilities in Children With Specific Language Impairment (SLI): Delayed or Deviant

Children with SLI exhibit a primary deficit in language, in the absence of documented neurological damage, hearing deficits, severe environmental deprivation, or mental retardation (Tomblin et al., 1997; Leonard, 2014). Bilingual children with SLI show deficits in both of their languages (Håkansson et al., 2003; Armon-Lotem and de Jong, 2015; Thordardottir, 2015). Similarly to the Weaker Language development, language development in children with SLI has been discussed in terms of delay and/or deviance (for an overview see Leonard, 2014). Delay suggests a typical pattern of acquisition, while deviance stands for disordered/atypical trajectory of language development.

Most studies addressing the delay-deviance debate have compared monolingual children with SLI to younger language-matched children with TLD (matched by MLU, vocabulary, grammar, general language skills). The Delay Hypothesis is reinforced by the findings that children with SLI have a late start, their language development is protracted, and their error

patterns are typical of younger children with TLD. For example, Rice et al. (1995), showed similarities in the acquisition of verbal morphology between monolingual children with SLI and younger children with TLD. For morpho-syntax, monolingual children with SLI were reported to perform similarly to younger language-matched controls (Stokes et al., 2006). The opposing view, the Deviance Hypothesis, has been advanced in studies reporting different error profiles in monolingual children with SLI and younger children with TLD. For instance, children with SLI have been shown to produce more bare stems compared to younger language-matched children in contexts, which require inflected forms (e.g., Bishop, 2014). Similarly, there are findings on morpho-syntactic abilities demonstrating distinct error profiles for children with SLI and younger language-matched controls (Briscoe et al., 2001; Riches, 2012). Moreover, it has been shown that language deficits in children with SLI may persist into adolescence (e.g., Conti-Ramsden et al., 2012), which would argue against the Delay Hypothesis or at least suggest that the initial delay becomes, in the long run, a deviance.

As for studies on bilingual children with SLI, the delay-deviance debate has not been addressed. Previous research shows similarities in linguistic profiles of monolingual and bilingual children with SLI, suggesting that disordered language development is similarly manifested irrespective of language status of a child (monolingual or bilingual). For example, Boerma et al. (2017) showed similarities in error profiles for participial affix use in Dutch among monolingual and bilingual children with SLI. Subject-verb agreement in German was reported to be similarly difficult for monolingual and bilingual children with SLI (Rothweiler et al., 2017). Russian-Hebrew speaking bilingual children with SLI were found to have difficulties with wh-questions and relative clauses (Meir et al., 2016) similarly to monolingual Hebrew speaking children with SLI (e.g., Friedmann and Novogrodsky, 2004; Novogrodsky and Friedmann, 2006). Similarly, monolingual and bilingual children with SLI showed difficulties with complex structures in German: wh-questions, relative clauses, embedding and finite complement clauses (Abed Ibrahim and Hamann, 2017; Hamann and Abed Ibrahim, 2017). In the same vein, monolingual and bilingual French speaking children with SLI were reported to have similar difficulties with morphology and syntax (Fleckstein et al., 2018).

To recap, the delay-deviance debate regarding language acquisition in children with SLI is still open. On the one hand, there are findings showing that children with SLI do not differ from younger language-matched controls which brings support to the claim that SLI is a delay. Conversely, there are studies showing quantitative and qualitative differences between children with SLI and younger children with TLD arguing for the Deviance Hypothesis.

## The Current Study

The present study has two aims. First, it attempts to advance our knowledge on the Weaker Language of unbalanced bilingual children with TLD. Furthermore, the study aims to bring new evidence for the delay-versus-deviance debate for the Weaker Language development in unbalanced bilingual children with

TLD and for language acquisition in children with SLI. The following research questions are addressed in the study:

- (1) To what extent there are quantitative and qualitative differences between unbalanced bilingual children with TLD in the Weaker Language and BB, on the one hand, and unbalanced bilinguals in the Dominant Language, on the other hand.
- (2) To what extent there are quantitative and qualitative differences between bilinguals with atypical language development (i.e., SLI) and unbalanced bilinguals with TLD in the Weaker Language.

Error profiles across different bilingual groups are expected to shed light on the nature of the Weaker Language development. It is hypothesized that similarities in error profiles of unbalanced bilinguals in the Weaker Language and BB and bilinguals in the Dominant Language would point at commonalities of language development in bilinguals with TLD irrespective of their dominance status. Differences in error profiles are hypothesized to signify different developmental patterns in the Weaker Language as compared to BB and bilinguals in the Dominant Language.

As for language development in children with SLI, it is hypothesized that similar linguistic profiles of unbalanced bilinguals in the Weaker Language and bilinguals with SLI would point at typical patterns of bilingual acquisition in the two populations favoring the Delay Hypothesis. By contrast, differences between bilinguals with SLI and unbalanced bilinguals in the Weaker Language would point at a disorder, rather than a delay, in children with SLI.

## MATERIALS AND METHODS

### Participants

For the purposes of the current study, one hundred and nineteen children aged 5;5–6;5 were drawn from a larger pool of participants (Meir, 2017). All bilingual children were recruited from regular and language preschools with SL-Hebrew as the language of instruction. All bilingual children were born in Israel to Russian-speaking families and were exposed to Russian from birth and had at least 12 months of exposure to SL-Hebrew.

Four groups of bilinguals were compared: three groups of bilinguals with TLD [unbalanced bilinguals with the Weaker HL-Russian and the Dominant SL-Hebrew (HL-weak:  $n = 39$ ), unbalanced bilinguals with the Weaker SL-Hebrew and the Dominant HL-Russian (SL-weak:  $n = 19$ ); balanced bilinguals (BB:  $n = 38$ ) and a group of bilingual children SLI (biSLI:  $n = 23$ )]. All children were tested on non-verbal IQ using Raven's colored progressive matrices non-verbal IQ test (Raven, 1998).

Language dominance in the current study was determined by language proficiency scores in both languages, following previous research (Paradis et al., 2003; Bedore et al., 2012; Iluz-Cohen and Armon-Lotem, 2013; Lust et al., 2016). In HL-Russian, language proficiency was measured using the *Russian*

*Language Proficiency Test for Multilingual Children* (Gagarina et al., 2010). The Russian proficiency test is comprised of a battery of expressive (noun/verb naming, production of case, and verb inflections) and receptive (comprehension of grammatical constructions, receptive vocabulary) subtests. In SL-Hebrew, language proficiency was tested using the *Goralnik Screening Test for Hebrew* (Goralnik, 1995). The Hebrew proficiency measure includes subtests for expressive vocabulary, sentence repetition, sentence comprehension, expression, pronunciation, and story-telling. Since proficiency measures in Russian and Hebrew were not parallel, provisional bilingual cut-off points were used (Altman et al., 2016), rather than subtracting scores in HL and SL.

Children with TLD were identified if there were no prior parental concern about their language development and scored within the bilingual norm in at least one of their languages (HL or SL). Children with TLD were assigned to the group of BB if they scored above  $-1.25$  SD in both of their languages. Unbalanced bilinguals with TLD were identified if they showed discrepancies in the proficiency scores. Children who scored below  $-1.25$  SD in HL-Russian, but above the cut-off point of  $-1.25$  SD in SL-Hebrew were labeled as unbalanced bilinguals with the Weaker Russian (HL-weak). Children who scored below  $-1.25$  SD in SL-Hebrew, but above the cut-off point of  $-1.25$  SD in HL-Russian were labeled as unbalanced bilinguals with the Weaker Hebrew (SL-weak).

Bilingual children with SLI (biSLI) were identified if they scored below  $-1.25$  SD in both languages using bilingual norms and had parent/teacher reported history of SLI/concerns about their language milestones or an evaluation by a certified SLP.

**Table 1** presents background information which was collected using a short version of the BIPAQ parental questionnaire (Abutbul-Oz et al., 2012). A one-way ANOVA showed that the four groups were matched for age [ $F(3,115) = 0.80$ ,  $p = 0.49$ ], socio-economic status as measured by maternal education in years [ $F(3,111) = 0.89$ ,  $p = 0.45$ ] and non-verbal IQ [ $F(3,115) = 0.04$ ,  $p = 0.99$ ].

By definition, there were group differences in language proficiency scores in HL-Russian [ $F(3,115) = 79.93$ ,  $p < 0.001$ ,  $\eta^2 = 0.68$ ] and in SL-Hebrew [ $F(3,115) = 75.87$ ,  $p < 0.001$ ,  $\eta^2 = 0.66$ ] (see **Table 2**). Balanced and unbalanced bilinguals in the Dominant Language outperformed bilinguals in the Weaker

**TABLE 1 |** Background information (means and standard deviations) on the participants per group.

	BB ( $N = 38$ )	HL-weak ( $N = 39$ )	SL-weak ( $N = 19$ )	biSLI ( $N = 23$ )
Age (in months)	71 (3)	71 (2)	72 (2)	72 (4)
Mother's education (in years)	15 (3)	14 (3)	14 (3)	14 (3)
Age of SL onset (in months)	37 (16)	23 (24)	47 (15)	38 (15)
Length of exposure to SL	34 (16)	48 (23)	26 (15)	34 (16)
Non-verbal IQ	113 (12)	113 (12)	114 (18)	113 (11)

BB, Balanced bilinguals; HL-weak, bilinguals with the Weaker Russian and the Dominant Hebrew; SL-weak, bilinguals with the Weaker Hebrew and Dominant Russian; biSLI, bilinguals with SLI; SL, Societal Language.



**TABLE 2 |** Language proficiency scores per group.

	BB (N = 38)	HL-weak (N = 39)	SL-weak (N = 19)	biSLI (N = 23)
Proficiency in HL-Russian (raw score)	87 (7)	57 (15)	87 (7)	51 (15)
Proficiency in HL-Russian (Z-score score)	0.16 (0.77)	−3.34 (1.69)	0.11 (0.75)	−4.01 (1.67)
Proficiency in SL-Hebrew (raw score)	146 (11)	148 (13)	110 (15)	100 (22)
Proficiency in SL-Hebrew (Z-score score)	0.16 (0.75)	0.30 (0.83)	−2.24 (1.00)	−2.88 (1.46)
Vocabulary scores in HL-Russian (subtest of the proficiency in HL-Russian)	36 (5)	18 (8)	37 (5)	17 (8)
Vocabulary scores in SL-Hebrew (subtest of the proficiency in SL-Hebrew)	16 (5)	15 (5)	8 (3)	10 (4)
Parental rating of HL-Russian (0–4 scale)*	3.62 (0.55)	2.69 (0.86)	3.67 (0.49)	2.43 (0.68)
Parental rating of SL-Hebrew (0–4 scale)*	2.95 (0.88)	3.42 (0.55)	2.50 (0.79)	2.38 (0.80)

\*Information was missing for 1 child in the SL-weak group; 2 children in the biSLI group; 3 children in the HL-weak group and 1 child in the BB group.

BB, balanced bilinguals; HL-weak, bilinguals with the Weaker Russian and the Dominant Hebrew; SL-weak, bilinguals with the Weaker Hebrew and Dominant Russian; biSLI, bilinguals with SLI; HL, Heritage Language; SL, Societal Language.

Language and the biSLI group [in HL-Russian: (BB = SL-weak) > (HL-weak = biSLI); in Hebrew: (BB = HL-weak) > (SL-weak = biSLI)].

Similarly to language proficiency scores, there were group differences in AoO [ $F(3,115) = 7.82$ ,  $p < 0.001$ ,  $\eta^2 = 0.17$ ]: [(BB = biSLI = SL-weak) < HL-weak]. Since length of exposure to SL-Hebrew is computed by deducting the AoO from the chronological age, similarly to AoO differences, there were significant differences for LoE [ $F(3,115) = 7.27$ ,  $p < 0.001$ ,  $\eta^2 = 0.16$ ]: [(BB = biSLI = SL-weak) > HL-weak]. Previous studies have used exposure measures as a proxy of language dominance. This study also shows that exposure measures (AoO and LoE) are linked to unbalanced bilingualism: unbalanced bilinguals with HL-weak and with SL-weak differed on AoO and LoE.

Since many studies, determine language dominance based on the discrepancy in the vocabulary size, expressive vocabulary scores for this sample are reported in **Table 2**. There was a group effect for HL-Russian [ $F(3,115) = 79.21$ ,  $p < 0.001$ ,  $\eta^2 = 0.67$ ] and for SL-Hebrew [ $F(3,115) = 36.65$ ,  $p < 0.001$ ,  $\eta^2 = 0.49$ ]. Follow-up pair-wise comparisons using Tamhane-2 *post hoc* tests for unequal variances revealed the following differences for HL-Russian: (BB = SL-weak) > (HL-weak = biSLI). Bonferroni *post hoc* tests showed a similar picture for SL-Hebrew: (BB = HL-weak) > (SL-weak = biSLI).

Parental ratings of the child's language skills in HL-Russian and SL-Hebrew were also noted using a 4-point scale: 1(poor) – 4(very good) (see **Table 2**). Correlational analysis revealed significant correlations between parental ratings and proficiency scores for HL-Russian [ $r(112) = 0.72$ ,  $p < 0.001$ ] and SL-Hebrew [ $r(112) = 0.53$ ,  $p < 0.001$ ]. Similarly to the results for the proficiency scores, the analysis of parental ratings indicated that there were significant group differences in HL-Russian [ $F(3,108) = 22.72$ ,  $p < 0.001$ ,  $\eta^2 = 0.39$ ] and SL-Hebrew [ $F(3,108) = 10.54$ ,  $p < 0.001$ ,  $\eta^2 = 0.32$ ]. Parental ratings converged with the direct assessment measures for HL-Russian ((BB = SL-weak) > (HL-weak = biSLI)). In SL-Hebrew, the biSLI group and the SL-weak group obtained similar ratings ( $p = 1.00$ ). The biSLI group received significantly lower scores than the BB and the HL-weak groups ( $p = 0.045$ ,  $p < 0.001$ ; respectively). Interestingly, SL-Hebrew parental

ratings of the BB with TLD were similar to those of the SL-weak and the HL-weak ( $p = 0.26$ ,  $p = 0.06$ ; respectively). These findings indicate that parents of BB children underestimate their children's abilities in the SL. This can be explained by the fact that BB, who have good language skills in both of their languages, conduct their communication in the HL with the parents; and maybe the parents did not master the SL themselves and cannot evaluate their children's ability in the SL.

## Procedure and Materials

The study was approved by Bar-Ilan University's IRB and by the Israeli Ministry of Education. Prior to the study, parental written consent forms were secured. Before each session, child assent was obtained. Each participant was tested individually in a quiet room at preschools. Testing was performed by native speakers of each language.

Sentence Repetition (SRep) tasks in Russian and in Hebrew were administered in two separate sessions, on different days. The order of language sessions (HL-Russian first, SL-Hebrew first) was counter-balanced. The experimental tasks were pre-recorded by native speakers of Russian and Hebrew for the consistency of presentation and were presented via a power-point presentation using earphones. The participants were instructed to repeat the stimuli orally verbatim. Practice items preceded the experimental items to ensure that the child understood the task.

The SRep tasks in Russian (Meir and Armon-Lotem, 2015) and in Hebrew (Meir et al., 2016) were based on LITMUS-SRep (Marinis and Armon-Lotem, 2015) developed within COST Action IS0804<sup>1</sup> and contained 56 sentences in each language (see **Tables A1, A2 in Appendix A**). The Russian and the Hebrew tasks elicit SVO sentences, biclausal sentences with coordination and subordination, object and oblique questions, object relatives and conditionals (real and unreal). The Russian SRep task additionally includes simple SOV and OVS sentences and subject relatives. The Hebrew SRep task additionally includes simple VSO sentences, oblique relative clauses and biclausal sentences with phrasal conjunctions. Following Marinis and Armon-Lotem

<sup>1</sup>www.bi-sli.org



(2015), the children's repetitions of the sentences were scored as correct if target structures were correctly reproduced. This scoring method enables to assess morpho-syntactic abilities of bilingual children without penalizing them for vocabulary errors. The proportion of correctly repeated structure out of 56 was calculated. Lexical substitutions were scored as correct (e.g., brother/boy, soup/food).

Furthermore, morphological accuracy was noted. Russian and Hebrew bilingualism offers an excellent opportunity to examine cross-linguistic influence, since the two languages vary in their selection of grammatical categories and vary in their mapping. For example, definiteness has an overt realization in Hebrew but not in Russian; aspect is realized in Russian but not in Hebrew. [ACC] case is realized in both languages, yet [ACC] case is differently mapped onto lexical categories in the two languages: in Russian [ACC] case is mapped onto nominal inflections, while in Hebrew [ACC] case is realized with the dedicated [ACC] marker *et* before [DEF] nouns. In Russian and in Hebrew verbal inflections mark categories of [Person], [Number], and [Gender].

A comparison of these morphological markings enables a fine-grained linguistic analysis in addressing directionality of cross-linguistic influence in bilingual children. The proportion of errors out of the total elicited items was calculated for each grammatical category. For example, in Russian and in Hebrew, verbal errors were analyzed in sentences in which verbs and overt subjects were produced. Sentences with null subjects were not included in the analysis. Erroneous use of [Person], [Number], [Gender] was noted: *ha-imahot \*SHOTIM qafe* 'the mothers.PL.FEM drink.PL.MASC coffee'; *mama \*POZVONIL* 'mother called.MASC'. Omissions of the definite marker *ha-* were noted only if the noun was produced: *imahot* 'mothers' instead of the targeted DP *ha-imahot* 'DEF mothers.' In Russian, erroneous use of the imperfective aspect marking was noted only on the elicited verbs: *tjotja \*MYLA posudu* 'aunt washed.IMPERF dishes' instead of *tjotja po-myla posudu* 'aunt washed.PERF dishes.' The same coding method was applied for coding [ACC] case errors on Russian nouns.

Furthermore, detailed error patterns analysis for each structure separately was conducted in order to shed light on grammatical representations in bilingual children (for more details on the analysis see Meir et al., 2016).

## Statistical Analysis

The data analysis was carried out using SPSS Statistics Version 18.0. First, group differences for global SRep scores and performance in each structure in Russian and Hebrew were analyzed with one-way ANOVAs with group (HL-weak, SL-weak, BB, biSLI) as an independent variable. Further pair-wise comparisons were conducted using Bonferroni *post hoc* tests for equal variance or Tamhane-2 *post hoc* tests for unequal variance with an adjusted alpha-level for multiple comparisons. The equality of variance was determined using the Levene's test.

To assess group differences on morphological markings, Kruskal-Wallis tests were applied with Mann-Whitney *U* tests as follow-ups for pair-wise comparisons. Finally, error profiles of

unbalanced bilinguals in the Weaker Language were compared to those of bilingual children with SLI using Mann-Whitney tests.

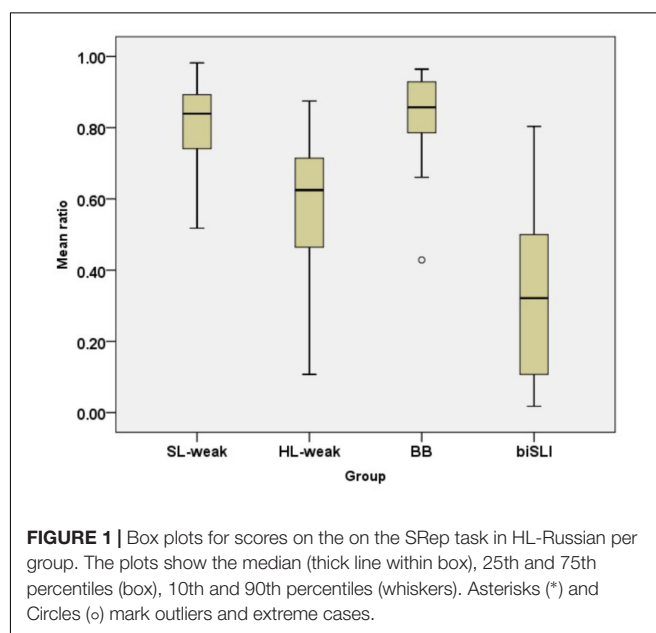
## RESULTS

### Findings for HL-Russian

#### Quantitative Comparison of the Four Bilingual Groups in HL-Russian

Figure 1 presents the performance on the SRep task in HL-Russian for the four bilingual groups. The analysis using a one-way ANOVA with children's scores on the SRep task in Russian as a dependent variable and group (HL-weak, SL-weak, BB, biSLI) as an independent variable showed a significant effect of group [ $F(3,112) = 51.83, p < 0.001, \eta^2 = 0.68$ ]. Pair-wise comparisons using Tamhane-2 *post hoc* tests showed that the HL-weak group scored lower than BB and unbalanced bilinguals in the Dominant Language, yet higher the biSLI group: (BB = SL-weak) > HL-weak > biSLI (all *p*-values at  $p < 0.001$ ).

Subsequently, the four groups were compared on 11 structures: group differences were detected for each structure (see Table 3). As determined by Tamhane-2 *post hoc* tests, the BB and the SL-weak groups scored similarly on all the structures. As for the HL-weak group, the comparison of their scores to the BB group showed a disadvantage on 9 out of 11 structures; no differences were found on biclausal sentences with coordination and subordination. Similarly, the HL-weak group scored lower than the SL-weak group on 8 out of 11 structures; no differences were found only for biclausal sentences with coordination and subordination and OVS sentences. As for the comparison of the HL-weak and the biSLI the analysis showed that the HL-weak group outperformed the biSLI group on 7 out of 11 structures. There were no significant differences between the two groups on four syntactic structures: biclausal sentences



**TABLE 3 |** Mean (SD) proportion of accuracy on 11 target structures in HL-Russian per group.

	BB (N = 38)	HL-weak (N = 37)	SL-weak (N = 38)	biSLI (N = 22)	F-results	$\eta^2$ -value
SVO	0.98 (0.08)	0.84 (0.18)	0.99 (0.04)	0.64 (0.29)	21.31**	0.36
SOV	0.72 (0.27)	0.43 (0.26)	0.58 (0.21)	0.19 (0.20)	23.79**	0.39
OVS	0.92 (0.17)	0.77 (0.26)	0.93 (0.11)	0.40 (0.33)	27.54**	0.42
Biclausal sentences with coordination	0.99 (0.06)	0.84 (0.24)	0.95 (0.10)	0.59 (0.40)	14.95**	0.29
Biclausal sentences with subordination	0.97 (0.08)	0.83 (0.31)	0.97 (0.08)	0.40 (0.38)	28.95**	0.44
Object questions	0.88 (0.18)	0.55 (0.29)	0.80 (0.21)	0.38 (0.32)	23.24**	0.38
Oblique questions	0.95 (0.12)	0.63 (0.33)	0.97 (0.08)	0.22 (0.29)	54.27**	0.59
Subject relatives	0.80 (0.18)	0.51 (0.28)	0.75 (0.22)	0.24 (0.31)	28.43**	0.43
Object relatives	0.57 (0.28)	0.23 (0.26)	0.57 (0.32)	0.13 (0.21)	20.07**	0.35
Real conditionals	0.96 (0.12)	0.70 (0.35)	0.93 (0.14)	0.55 (0.43)	12.66**	0.25
Unreal conditionals	0.61 (0.36)	0.16 (0.26)	0.63 (0.38)	0.03 (0.09)	28.87**	0.44

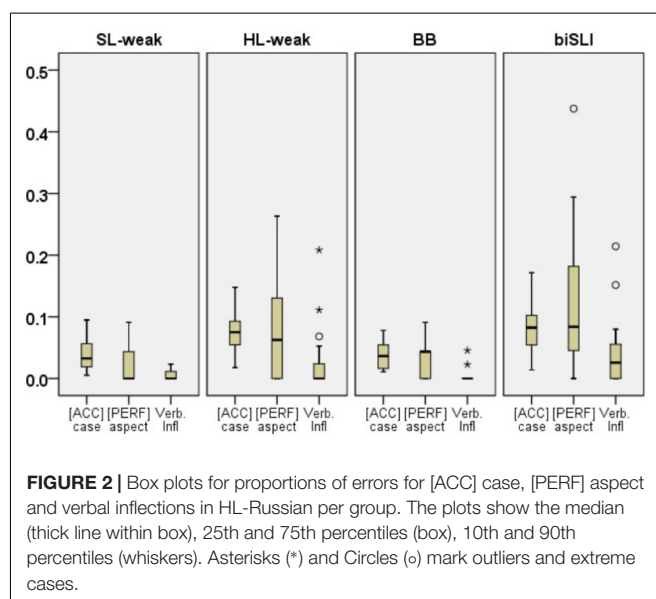
\*\*Significance at  $p < 0.001$ . BB, balanced bilinguals; HL-weak, bilinguals with the Weaker HL-Russian and the Dominant SL-Hebrew; SL-weak, bilinguals with the Weaker SL-Hebrew and Dominant HL-Russian; biSLI, bilinguals with SL; SVO, Subject-Verb-Object; SOV, Subject-Object-Verb; OVS, Object-Verb-Subject.

with coordination, object questions, object relatives and real conditionals.

Further analysis compared morphological accuracy in HL-Russian across the four groups (see **Figure 2**). The Kruskal-Wallis test showed a group effect for [ACC] case errors [ $\chi^2(3) = 39.88, p < 0.001$ ], [PERF] aspect [ $\chi^2(3) = 21.53, p < 0.001$ ] and verbal inflections [ $\chi^2(3) = 31.41, p < 0.001$ ]. Further pair-wise comparisons using Mann-Whitney  $U$  tests showed no differences for the BB and the SL-weak on [ACC] case ( $U = 56, p = 0.93$ ), and [PERF] aspect ( $U = 331, p = 0.58$ ), yet there were differences between the two groups on verbal inflections ( $U = 287, p = 0.03$ ) with the BB group being more accurate on verbal inflections. The HL-weak group was less accurate on [ACC] case and [PERF] aspect than the BB ( $U = 01, p < 0.001$ ;  $U = 427, p < 0.001$ , respectively) and the SL-weak groups ( $U = 137, p < 0.001$ ;  $U = 200, p = 0.01$ , respectively). Yet, on verbal inflections the HL-weak showed marginal differences with the SL-weak group ( $U = 265, p = 0.08$ ) and significant differences from the BB group ( $U = 432, p < 0.001$ ). The comparison of the HL-weak and the biSLI groups showed no significant differences between the two groups for [ACC] case errors ( $U = 317, p = 0.16$ ) and for [PERF] aspect errors ( $U = 354, p = 0.27$ ). Group differences were observed for verbal inflection errors ( $U = 291, p = 0.03$ ) with the biSLI group being less accurate.

### Comparison of Morpho-Syntactic Profiles in HL-Russian

Subsequently, error profiles of the four bilingual groups were investigated. No differences were detected between the BB group and the SL-weak. Despite quantitative differences between the BB and the HL-weak groups, error profiles of the two groups seem to overlap. The only pattern which differentiated the HL-weak group from the BB group was the substitution of the *wh*-pronoun (inflected for case, number, and gender) with the non-declinable complementizer 'čto' in subject and object relative clauses (both comparisons at  $p < 0.001$ ). This error might be attributed to the influence of Dominant-Hebrew, which uses non-declinable complementizer 'še' in subject and object relatives.



As for the comparison of the unbalanced bilinguals in their Weaker Language (HL-weak) and the biSLI, different error profiles emerged across several structures (see **Table 4**). The biSLI group produced more sentence fragments, omitted conjunctions and simplified structures (e.g., produced simple SVO sentences instead of targeted object questions, object relatives and subject relatives). Interestingly, the HL-weak group and the biSLI had similar accuracy scores on object relatives, yet error analysis showed that the underlying difficulties were of different natures [see Example (1)].

Children in the HL-weak group attempted to re-produce a complex structure [see Examples (1)]. Some HL-weak had difficulties with case inflections, producing both elements either in [NOM] or [ACC] (see 1a and 1b), some children substituted an inflected *wh*-pronoun with a non-declinable complementizer (the Hebrew *še* or the Russian *čto*) (see 1c and 1d). Conversely, children in the biSLI group turned

(1)	Target	<i>jeto</i>	<i>devočka</i>	<i>kotor-uju</i>	<i>narisovala</i>	<i>mama</i>
		this	girl.NOM	who-F.SG.ACC	drew.PERF	mother.NOM
		'This is the girl that the mother drew.'				
Responses of the children in the Weak-RUS group						
(a)		<i>jeto</i>	<i>devočka</i>	<i>kotor-<u>aja</u></i>	<i>risovala</i>	<i>mam-a</i>
		this	girl.NOM	who- <b>F.SG.NOM</b>	drew.IMPERF	mother.NOM
(b)		<i>jeto</i>	<i>devočka</i>	<i>kotor-<u>uju</u></i>	<i>risovala</i>	<i>mam-a</i>
		this	girl.NOM	who- <b>F.SG.ACC</b>	drew.IMPERF	mother.NOM
(c)		<i>jeto</i>	<i>devočka</i>	<b><u>še</u></b>	<i>narisovala</i>	<i>mam-a</i>
		this	girl.NOM	that (Hebrew complementizer)	drew.PERF	mother.NOM
(d)		<i>jeto</i>	<i>devočka</i>	<b><u>čto</u></b>	<i>narisovala</i>	<i>mam-<b>u</b></i>
		this	girl.NOM	that(complementizer)	drew.PERF	mother.ACC
(e)		<i>jeto</i>	<i>devočka</i>	<b><i>kak-oj</i></b>	<i>narisovala</i>	<i>mam-<b>u</b></i>
		this	girl.NOM	which- <b>M.SG.NOM</b> (wh-word)	drew.PERF	mother.ACC
Responses of the children in the biSLI group						
(f)		<i>jeto</i>	<i>devočka</i>	<i>narisovala</i>		
		this	girl.NOM	drew.PERF		
(g)			<i>devočka</i>	<i>narisovala</i>	<i>mam-<b>u</b></i>	
			girl.NOM	drew.PERF	mother.ACC	
(h)		<i>jeto</i>	<i>devočka</i>	<i>narisovala</i>	<i>mam-<b>a</b></i>	
		this	girl.NOM	drew.PERF	mother.NOM	

object relatives into simple SV or SVO sentences (see 1g-h).

To sum up, the results for HL-Russian have demonstrated quantitative differences between the Weaker Language of bilinguals (HL-weak) and BB and unbalanced bilinguals in the Dominant Language. The morphological accuracy of unbalanced bilinguals in the Weaker Language was lower than in BB and bilinguals in the Dominant Language. In the Weaker Language, the unbalanced bilinguals showed similar performance to the biSLI group for [ACC] case and [PERF] aspect (on the features that are differently configured in Russian and Hebrew). On verbal inflections, the HL-weak group scored lower than the BB and the SL-weak, yet HL-weak outperformed the biSLI. Even though the HL-weak and the BB/SL-weak groups showed quantitative differences, their error profiles

overlapped for most structures. The only error pattern which differentiated the two groups was wh-pronoun substitution with the complementizers in Russian or in Hebrew (e.g., 'čto'/'še'): this error pattern can be traced back to the influence of the Dominant-Hebrew.

Despite similar vocabulary scores in HL- Russian, the HL-weak group outperformed the biSLI group on the global SRep score and on a variety of structures. Both groups (HL-weak and biSLI) showed low accuracy on morphological categories that are differently configured in Russian and Hebrew (e.g., [ACC] case and [PERF] aspect). Importantly, error profiles of the HL-weak and the biSLI group were found to bear fundamental differences. While children in the biSLI simplified complex structures, unbalanced bilinguals in the Weaker Language opted for complex structures.

**TABLE 4 |** Proportions of most prominent syntactic error patterns observed on SRep in HL-Russian (HL-weak vs. biSLI).

Target structure	Error type	HL-weak Mean	biSLI Mean	U-values (Mann-Whitney test)	p-value (Mann-Whitney test)
Biclausal with coordination	Sentence fragment	0.03	0.13	343	$p = 0.006$
Biclausal with subordination	Conjunction omission	0.04	0.27	242	$p < 0.001$
Real conditional	Conjunction omission	0.01	0.14	334	$p < 0.001$
Unreal conditional	Conjunction omission	0.01	0.08	352	$p = 0.035$
Oblique question	Preposition omission	0.08	0.31	242	$p = 0.001$
Object question (OQ)	OQ into SVO	0.04	0.16	291	$p = 0.007$
Object relative (OR)	OR into SVO	0.04	0.33	161	$p < 0.001$
	OR into wh-question	0.00	0.02	351	$p = 0.006$
	Case error	0.34	0.12	224	$p = 0.001$
Subject relative (SR)	SR into SVO	0.06	0.40	201	$p < 0.001$
	SR into wh-question	0.00	0.04	342	$p = 0.011$

HL-weak, bilinguals with the Weaker HL-Russian and the Dominant SL-Hebrew; biSLI, bilinguals with SLI; SR, subject relatives; OR, object relatives.

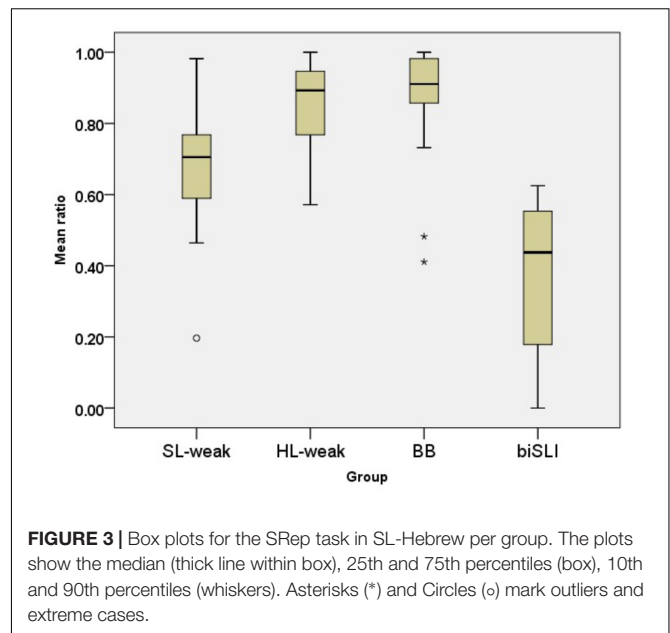
## Findings for SL-Hebrew

### Quantitative Comparison of the Four Bilingual Groups in SL-Hebrew

Turning to the SL-Hebrew data, **Figure 3** presents the performance on the SRep task in SL-Hebrew. The analysis using a one-way ANOVA with children's scores on the SRep task in Hebrew as a dependent variable and group (HL-weak, SL-weak, BB, biSLI) as an independent variable showed a significant effect of group [ $F(3,112) = 64.69, p < 0.001, \eta^2 = 0.63$ ]. Follow-up pair-wise comparisons using Tamhane-2 *post hoc* showed that, similarly to the Russian data, unbalanced bilinguals with TLD in their Weaker Language scored lower than BB and bilinguals in the Dominant Language, yet higher than bilinguals with SLI [(BB = HL-weak) > SL-weak > biSLI].

Further analyses compared the performance of the four groups across the 11 structures of the Hebrew SRep Task (see **Table 5**). There was a group effect for all the structures. Follow-up pair-wise comparisons using Tamhane-2 *post hoc* tests showed that the BB and the HL-weak groups scored similarly across all the structures. The SL-weak scored lower than the BB group on 6 of 11 structures: no differences were observed for 5 structures (SVO, biclausal with coordination, biclausal with subordination, oblique questions, and object relatives). Similarly, the SL-weak scored lower than the HL-weak, i.e., bilinguals with the Dominant SL-Hebrew and the Weaker HL-Russian, on 7 out of 11 structures: no differences were detected for 4 structures (SVO, biclausal with coordination, biclausal with subordination, and object relatives). The comparison of the SL-weak and the biSLI group showed that the unbalanced bilinguals in the Weaker SL-Hebrew outperformed the biSLI group on 9 out of the 11 tested structures, no differences between the two groups were found for the unreal conditionals and biclausal sentences with coordination.

Subsequently, morphological accuracy (proportion of errors for the [DEF] marker *ha-* and verbal inflections) was compared



across the four groups (see **Figure 4**). The Kruskal–Wallis test showed a group effect for [DEF] marking errors [ $\chi^2(3) = 42.928, p < 0.001$ ] and verbal inflections [ $\chi^2(3) = 23.16, p < 0.001$ ].

A Mann–Whitney *U* test indicated no significant differences between the BB and the HL-weak group for both morphemes ([DEF] marker:  $U = 647, p = 0.33$ , verbal inflections:  $U = 660, p = 0.29$ ). The SL-weak group showed lower accuracy than the BB ( $U = 158, p < 0.001$ ) and the HL-weak on [DEF] marker ( $U = 140, p < 0.001$ ). On verbal inflections the SL-weak also showed lower performance than the HL-weak ( $U = 228, p = 0.01$ ) and marginally lower than the BB group ( $U = 256, p = 0.08$ ). As for the SL-weak and the biSLI group comparisons, the analysis showed significantly more omission of the [DEF] marker *ha-* in

**TABLE 5 |** Mean (SD) proportion of accuracy on 11 target structures in SL- Hebrew per group.

	BB (N = 38)	HL-weak (N = 39)	SL-weak (N = 18)	biSLI (N = 22)	F-results	$\eta^2$ -value
SVO	0.98 (0.04)	0.96 (0.07)	0.94 (0.13)	0.70 (0.26)	25.59**	0.41
Biclausal sentences with coordination)	0.90 (0.15)	0.93 (0.17)	0.74 (0.33)	0.48 (0.41)	17.73**	0.32
Biclausal sentences with subordination	0.97 (0.10)	0.96 (0.12)	0.83 (0.26)	0.47 (0.40)	29.15**	0.44
Object questions	0.83 (0.25)	0.79 (0.25)	0.54 (0.29)	0.27 (0.27)	26.58**	0.41
Oblique questions	0.91 (0.18)	0.83 (0.25)	0.67 (0.33)	0.14 (0.24)	54.41**	0.59
Object relatives	0.93 (0.16)	0.93 (0.15)	0.85 (0.24)	0.45 (0.39)	23.88**	0.39
Oblique relatives	0.94 (0.15)	0.96 (0.11)	0.72 (0.30)	0.35 (0.37)	42.56**	0.53
VSO	0.88 (0.18)	0.83 (0.20)	0.69 (0.21)	0.41 (0.30)	24.27**	0.39
Real conditionals	0.95 (0.10)	0.91 (0.15)	0.74 (0.25)	0.45 (0.33)	33.54**	0.47
Unreal conditionals	0.76 (0.24)	0.68 (0.31)	0.35 (0.34)	0.16 (0.27)	25.57**	0.40
Biclausal sentences with phrasal conjunctions	0.72 (0.28)	0.73 (0.27)	0.37 (0.31)	0.09 (0.12)	37.10**	0.50

\*\*Significance at  $p < 0.001$ .

BB, Balanced bilinguals; HL-weak, bilinguals with the Weaker HL-Russian and the Dominant SL-Hebrew; SL-weak, bilinguals with the Weaker SL-Hebrew and Dominant HL-Russian; biSLI, bilinguals with SLI; SVO, Subject–Verb–Object; VSO, Verb–Subject–Object.



the biSLI group ( $U = 100, p = 0.01$ ), and no differences between the two groups for the verbal inflections ( $U = 134, p = 0.12$ ).

### Comparison of Error Profiles in SL-Hebrew

Error pattern analysis showed that the BB group and the HL-weak group, i.e., unbalanced bilinguals who are dominant in SL-Hebrew, showed identical error profiles in their SL-Hebrew. Moreover, no differences were detected for error profiles of the SL-weak group and BB. However, differences in error profiles emerged between the SL-weak and the biSLI groups. **Table 6** presents the most prominent error patterns for the two groups (biSLI vs. SL-weak). Similarly to the Russian data, in SL-Hebrew the biSLI group turned complex sentences into simpler sentences (e.g., object questions were turned into simple SVO sentences) and had significantly more preposition and conjunction omissions.

As demonstrated in (2), the SL-weak group reproduced object relatives (see 2a–b) while children in the biSLI group simplified relative clauses and produced simple SVO sentences or subject relatives (see examples 2d–e).

- (2) Target    *zot    ha- yalda    še    ha- iša    niška*  
                  this   DEF- girl   that   DEF- woman   kissed  
                  ‘This is the girl that.’

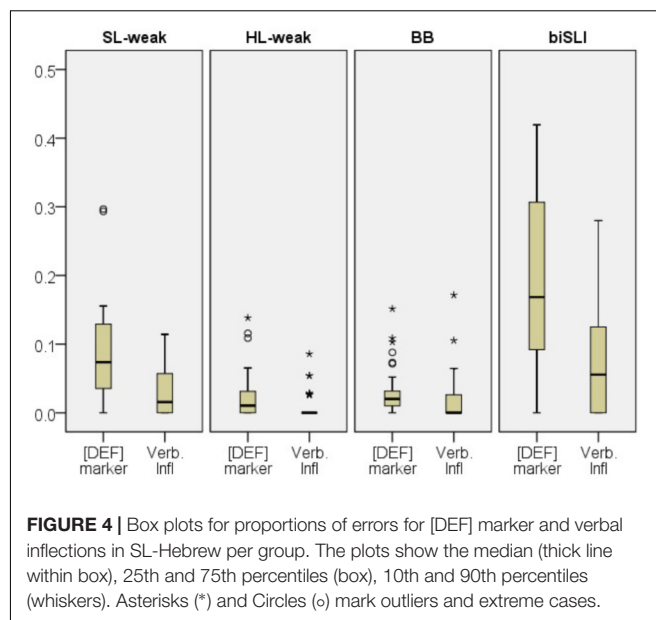
Responses of the children in the Weak-RUS group

- (a)    *zot    \_\_ yalda    še    \_\_ iša    niška*  
          this   \_\_ girl   that   \_\_ woman   kissed  
 (b)    *zot    ha- yalda    še    \_\_ yalda    niška*  
          this   DEF- girl   that   \_\_ girl   kissed

Responses of the children in the biSLI group

- (a)    *zot    ha- yalda    niška    \_\_ iša*  
          this   DEF- girl   kissed   \_\_ woman  
 (b)    *zot    ha- yalda    še    niška    \_\_ iša*  
          this   DEF- girl   that   kissed   \_\_ woman

The findings for SL-Hebrew converge with the results for HL-Russian: unbalanced bilinguals with TLD in the Weaker Language (i.e., SL-weak) differ from BB and bilinguals in the Dominant Language only quantitatively, while error profiles of all bilinguals with TLD bear a striking resemblance. Yet, unbalanced bilinguals in the Weaker Language show quantitative and qualitative differences from the biSLI group. Unbalanced bilinguals outperform the biSLI group on a number of measures and show different profiles from the biSLI group. While the former succeeded in reproducing complex structures despite their limited vocabulary, the latter simplified complex structures.



**FIGURE 4 |** Box plots for proportions of errors for [DEF] marker and verbal inflections in SL-Hebrew per group. The plots show the median (thick line within box), 25th and 75th percentiles (box), 10th and 90th percentiles (whiskers). Asterisks (\*) and Circles (o) mark outliers and extreme cases.

## DISCUSSION

The current study was devised to determine to what extent morpho-syntactic abilities of bilingual children with TLD in the Weaker Language differ from those of BB and bilinguals in the Dominant Language, on the one hand, and bilingual children with SLI, on the other hand. This study attempted to add to the on-going debate on the nature of grammatical representations and developmental trajectories among unbalanced bilinguals in the Weaker Language and bilingual children with SLI. To achieve this goal, different bilingual patterns of acquisition were investigated. This study addressed the delay-deviance hypothesis in bilingual children rather than comparing monolingual and bilingual trajectories of acquisition.

### The Weaker Language and the Balanced/Dominant Language of Bilinguals

The first research question of the current study aimed to explore morpho-syntactic manifestations in the Weaker Language of unbalanced bilinguals with TLD as compared to BB and bilinguals in the Dominant Language. Numerous studies have demonstrated quantitative differences between the Weaker

**TABLE 6 |** Proportions of most prominent syntactic error patterns observed on SRep in SL-Hebrew (SL-weak vs. biSLI).

Target structure	Error type	SL-weak Mean	biSLI Mean	U-values (Mann-Whitney test)	p-value (Mann-Whitney test)
Oblique question (OQ)	OQ into subject question	0.01	0.18	82	$p = 0.002$
	Preposition omission	0.17	0.38	112	$p = 0.030$
Object relatives (OR)	OR into SVO	0.07	0.23	115	$p = 0.037$
Advanced conjunctions	Conjunction omission	0.12	0.30	105	$p = 0.017$

SL-weak, bilinguals with the Weaker Hebrew and Dominant Russian; biSLI, bilingual children with SLI; OQ, Object Question; OR, Object Relative.

Language of unbalanced bilinguals and BB and bilinguals in the Dominant Language (e.g., Schlyter, 1994; Jisa, 2000; Bernardini and Schlyter, 2004; Gutiérrez-Clellen et al., 2006; Bedore et al., 2011; Hoff et al., 2012). Yet, previous studies brought conflicting evidence with respect to qualitative differences between bilinguals in the Weaker Language and monolinguals and BB and bilinguals in the Dominant Language. Some studies have shown that the acquisition patterns in the Weaker Language are similar to the ones of BB and bilinguals in the Dominant Language and even monolinguals (e.g., Müller and Kupisch, 2003; Bernardini and Schlyter, 2004; Antonova Ünlü and Li, 2016, 2017, 2018). Alternatively, the Deviance Hypothesis was supported by findings indicating that grammars of unbalanced bilinguals in their Weaker Language differ qualitatively from the monolingual baseline grammars (e.g., Yip and Matthews, 2000; Argyri and Sorace, 2007; Ringblom, 2012; Janssen, 2016; Meir et al., 2017). Previous studies have brought convincing evidence that the two linguistic systems of a bilingual person are susceptible to bi-directional cross-linguistic influence: influence from HL onto SL and from SL onto HL (e.g., Ge et al., 2017; Hervé and Serratrice, 2017; Meir et al., 2017). This study has focused on different bilingual outcomes rather than comparing bilingual performance to a monolingual “golden standard.”

The results of the current study reiterate previous findings showing that the Weaker Language of unbalanced bilinguals is quantitatively poorer. Unbalanced bilinguals in the Weaker Language have smaller vocabularies and are less accurate on a variety of morpho-syntactic structures as compared to BB and bilinguals in the Dominant Language. Moreover, unbalanced bilinguals in the Weaker Language have more pronounced difficulties with morphology as compared to BB and bilinguals in the Dominant Language. For example, in HL-Russian unbalanced bilinguals with the Weaker Russian showed lower accuracy for [ACC] case marking and for [PERF] aspect marking in comparison with BB and bilinguals in the Dominant Language.

However, despite quantitative differences, error profiles of unbalanced bilinguals in the Weaker Language bore resemblance to the ones of BB and bilinguals in the Dominant Language, extending the finding by Müller and Kupisch (2003) and Bernardini and Schlyter (2004) to two different patterns of unbalanced bilingual acquisition: for the Weaker HL and the initial Weaker SL. Despite lower scores in HL-Russian on various syntactic structures, error patterns in unbalanced bilinguals with the Weaker HL-Russian were similar to those of BB and bilinguals who are Dominant in HL-Russian. That is all bilinguals with TLD had difficulties with case inflections. Previously, case inflectional morphology has been reported to pose difficulties to bilingual children who acquire Russian as their HL and SL that does not mark cases with inflections (Turian and Altenberg, 1991; Peeters-Podgaevskaja, 2008; Gagarina, 2011; Schwartz and Minkov, 2014; Janssen et al., 2015; Meir and Armon-Lotem, 2015). In HL-Russian, unbalanced bilinguals with the Weaker HL-Russian and the Dominant SL-Hebrew substituted the declinable *wh*-pronoun (marked for case, gender, and number) with the non-declinable complementizer ‘*čto*’ in subject and object relative clauses. This substitution can be easily attributed to the influence of the Dominant SL-Hebrew, since Hebrew utilizes non-declinable

caseless complementizers ‘*še*’ in relative clauses. Moreover, such an error pattern has been previously reported for a child with the Weaker Russian and the Dominant Swedish: the child used a Swedish uninflected complementizers *som* in relative clauses instead of a Russian declinable *wh*-pronoun [e.g., *eto ja som sdelał eto* ‘this I who did this’ (see Dobrova and Ringblom, 2018)].

Similarly, in SL-Hebrew, there were quantitative differences between unbalanced bilinguals with the Weaker SL-Hebrew and BB and bilinguals with the Dominant SL-Hebrew. Bilinguals with the Weaker SL-Hebrew showed lower performance for the global score on the Hebrew SRep task. They also showed lower levels of accuracy on nearly half of the morpho-syntactic structures as compared to BB and bilinguals with the Dominant SL-Hebrew. Bilinguals with the Weaker SL-Hebrew were less accurate on morphological markings of definiteness and verbal inflections as compared to BB and bilinguals with the Dominant SL-Hebrew.

Yet, importantly, the analysis of error profiles of unbalanced bilinguals with the Weaker SL-Hebrew and BB and bilinguals in the Dominant Language showed that there are no differences between the groups. The results indicate that bilinguals with TLD who have mastered complex constructions in their HL seem to draw on their existing linguistic knowledge to produce complex structures in the SL, albeit their poor lexicons and morphology in the Weaker Language.

Future research exploring grammatical representations of unbalanced bilinguals with TLD should explore other language pairs in order to deepen our understanding on how typological differences affect grammatical representations in the Weaker Language under the influence of the Dominant Language. Research on unbalanced bilingualism should be extended to school-age children. This line of research would enable us to evaluate how grammatical representations of BB and unbalanced bilinguals with the Weaker HL, who get extensive exposure and acquire literacy skills in their SL, change over time.

## Delay-Deviance Debate: Unbalanced TLD in the Weaker Language and Atypical Language Development

The second research question of the study aimed to contribute to the delay-versus-deviance debate on language acquisition patterns of unbalanced bilinguals with TLD in the Weaker Language and bilinguals with SLI. Language development in unbalanced bilinguals in the Weaker Language is not expected to be disordered, while it may be delayed and/or influenced by the Dominant Language. Thus, it was hypothesized that similarities in the profiles of the Weaker Languages of bilinguals with TLD and bilinguals with SLI would point at similar morpho-syntactic representations in the two populations. Qualitative differences in error profiles of unbalanced bilinguals with TLD in the Weaker Language and bilinguals with SLI were predicted to support a deviant language acquisition pattern in children with SLI.

Despite similar vocabulary sizes of unbalanced bilinguals in the Weaker Language and bilinguals with SLI, the former showed higher scores on SRep tasks. More importantly, error profiles

of the two populations were found to be different. Whereas unbalanced bilinguals in the Weaker Language opted for complex structures relying on the available resources from the Dominant Language, bilinguals with SLI opted for simplified structures. Bilinguals with SLI produced simple SVO sentences instead of targeted object questions, object relatives and subject relatives. This has been also found for monolingual children with SLI (e.g., Novogrodsky and Friedmann, 2006). Previous research has pointed at similarities in linguistic profiles of monolingual and bilingual children with SLI, suggesting that disordered language development is similarly manifested irrespective of language status of a child (monolingual or bilingual) (see Meir et al., 2016; Abed Ibrahim and Hamann, 2017; Boerma et al., 2017; Hamann and Abed Ibrahim, 2017; Rothweiler et al., 2017). Importantly, this has been confirmed for both languages of bilinguals with SLI (HL-Russian and SL-Hebrew).

The current study convincingly shows that grammatical representations in unbalanced bilinguals in the Weaker Language and bilinguals with SLI differ. The results for bilinguals with SLI couple with the literature on monolingual children with SLI suggesting that language profiles of children with SLI show a deviant pattern of acquisition (e.g., Briscoe et al., 2001; Conti-Ramsden et al., 2012; Riches, 2012; Bishop, 2014).

## CONCLUSION

The current study assessed grammatical representations of unbalanced bilinguals in the Weaker Language in the group of Russian-Hebrew speaking pre-school children. The findings indicate that grammatical representations of unbalanced bilinguals (either HL or SL) are qualitatively similar to the ones of BB and bilinguals in the Dominant Language, albeit their performance is quantitatively disadvantaged. The Weaker Language of bilinguals is characterized by an increased number of morphological errors, especially when the two languages (HL and SL) show differences in the selection and mapping of morpho-syntactic categories. The findings indicate that mastering rich morphology in the context of reduced input is challenging. Yet, despite limited vocabulary size and limited arsenal of morphological markings, unbalanced bilinguals attempt to derive complex structures in the Weaker Language recruiting resources from their Dominant Language.

## REFERENCES

- Abed Ibrahim, L., and Hamann, C. (2017). "Bilingual Arabic-German and Turkish-German children with and without specific language impairment: comparing performance in sentence and nonword repetition tasks," in *Proceedings of the 41st Annual Boston University Conference on Language Development, BUCLD 41*, eds M. LaMendola and J. Scott (Somerville, MA: Cascadia Press), 1–17.
- Abutbul-Oz, H., Armon-Lotem, S., and Walters, J. (2012). *Bilingual Parents Questionnaire (BIPAQ)*. Ramat Gan: Bar-Ilan University.
- Altman, C., Harel, E., Meir, N., Iluz-Cohen, P., Walters, J., and Armon-Lotem, S. (2016). The Goralnik Screening Test for Hebrew: bilinguals norms. *Paper*

The comparison of morpho-syntactic abilities of unbalanced bilinguals in the Weaker Language to those of bilinguals with SLI has demonstrated that the disordered pattern of acquisition is different from that of the Weaker Language development in unbalanced bilinguals with TLD. Whereas unbalanced bilinguals in the Weaker Language attempt to produce complex structures, relying on the available resources from the Dominant Language; bilinguals with SLI simplified structures.

## ETHICS STATEMENT

This study was approved by Bar Ilan University review board as well as by the Israeli Ministry of Education. Parents provided informed written consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

NM developed the research questions, carried out the data analyses, and wrote the manuscript.

## FUNDING

The data collection was supported by The Israel Science Foundation (grants nos. 779/10 and 863/14) and the German Israel Foundation (grant no. 1113/2010). The writing of the manuscript was supported by The Israel Science Foundation (grant no. 1068/16).

## ACKNOWLEDGMENTS

The data were collected as part of my Ph.D. at the university of Bar-Ilan (Israel) under the supervision of Prof. Sharon Armon-Lotem to whom I would like to express my deep gratitude. I would like to thank Dr. Rama Novogrodsky, the two reviewers and the Editors for their most insightful comments and suggestions on previous versions of the manuscript. This paper benefited from the insights of the audience at BUCLD 41 and EUCLDIS Meeting 2016. Last but not least, I thank all the families that took part in this study.

*Presented at the 52nd Annual Conference of the Israeli Speech Hearing and Language Association, Jerusalem.*

- Antonijevic, S., Durham, R., and Chonghaile, I. (2017). Language performance of sequential bilinguals on an Irish and English sentence repetition task. *Linguist. Approaches Biling.* 7, 359–393. doi: 10.1075/lab.1502.6.ant
- Antonova Ünlü, E., and Li, W. (2016). Aspect acquisition in Russian as the weaker language: evidence from a Turkish-Russian child. *Int. J. Biling.* 20, 210–228. doi: 10.1177/1367006914552138
- Antonova Ünlü, E., and Li, W. (2017). The acquisition of the weaker language: evidence from the acquisition of Russian cases by a Turkish-Russian child. *Linguist. Approaches Biling.* doi: 10.1075/lab.16029.ant

- Antonova Ünlü, E., and Li, W. (2018). Examining the effect of reduced input on language development: the case of gender acquisition in Russian as a non-dominant and dispreferred language by a bilingual Turkish–Russian child. *Int. J. Biling.* 22, 215–233.
- Archibald, L. M., and Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *J. Speech Lang. Hear. Res.* 52, 899–914. doi: 10.1044/1092-4388(2009/08-0099)
- Argyri, E., and Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism* 10, 79–99. doi: 10.1017/S1366728906002835
- Armon-Lotem, S. (2014). Between L2 and SLI: inflections and prepositions in the Hebrew of bilingual children with TLD and monolingual children with SLI. *J. Child Lang.* 41, 3–33. doi: 10.1017/S0305000912000487
- Armon-Lotem, S., and de Jong, J. (2015). “Introduction,” in *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*, eds S. Armon-Lotem, J. de Jong, and N. Meir (Bristol: Multilingual Matters), 1–24. doi: 10.21832/9781783093137
- Bedore, L. M., Peña, E. D., Joyner, D., and Macken, C. (2011). Parent and teacher rating of bilingual language proficiency and language development concerns. *Int. J. Bilingual Educ. Biling.* 14, 489–511. doi: 10.1080/13670050.2010.529102
- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., et al. (2012). The measure matters: language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism* 15, 616–629. doi: 10.1017/S1366728912000090
- Bernardini, P. (2017). Weak interest in the weaker language. *Bilingualism* 20, 29–30. doi: 10.1017/S1366728916000341
- 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
- Bishop, D. (2014). *Uncommon Understanding (Classic Edition): Development and Disorders of Language Comprehension in Children*. Hove: Psychology Press.
- 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-L-16-0351
- Briscoe, J., Bishop, D. V. M., and Norbury, F. C. (2001). Phonological processing, language, and literacy: a comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *J. Child Psychol. Psychiatry Allied Discip.* 42, 329–340. doi: 10.1111/1469-7610.00726
- Conti-Ramsden, G., Botting, N., and Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *J. Child Psychol. Psychiatry* 42, 741–748. doi: 10.1111/1469-7610.00770
- 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)
- de 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, 331–358. doi: 10.1016/j.cortex.2013.11.006
- De Houwer, A. (1990). *The Acquisition of Two Languages from Birth: A Case Study*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511519789
- Dobrova, G. R., and Ringblom, N. (2018). Rechevoje razvitije rebjonka – simultannogo bilingva: dvojanaja sistema ili edinaja sistema “drugogo razvitija? Paper Presented at the Problemy Ontolingvistiki – 2018, Saint Petersburg.
- Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E., and Zebib, R. (2018). How to identify SLI in bilingual children: a study on sentence repetition in French. *Lang. Acquis.* 25, 85–101. doi: 10.1080/10489223.2016.1192635
- Friedmann, N., and Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development. *J. Child Lang.* 31, 661–681. doi: 10.1017/S0305000904006269
- Gagarina, N. (2011). “Acquisition and loss of L1 in a Russian–German bilingual child: a case study,” in *Monolingual and Bilingual Path to Language*, ed. S. Cjetlin (Moscow: Jazyki Slavyanskoy Kul'tury), 137–163.
- Gagarina, N., Klassert, A., and Topaj, N. (2010). “Russian language proficiency test for multilingual children,” in *ZAS Papers in Linguistics*, ed. M. André (Berlin: ZAS), 54.
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingualism* 12, 213–237. doi: 10.1017/S1366728909004015
- Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., et al. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Front. Psychol.* 5:11. doi: 10.3389/fpsyg.2014.00011
- Gavarró, A. (2017). A Sentence Repetition Task for Catalan-speaking typically-developing children and children with Specific Language Impairment. *Front. Psychol.* 8:1865. doi: 10.3389/fpsyg.2017.01865
- Ge, H., Matthews, S., Cheung, L. Y. L., and Yip, V. (2017). Bidirectional cross-linguistic influence in Cantonese–English bilingual children: the case of right-dislocation. *First Lang.* 37, 231–251. doi: 10.1177/0142723716687955
- Goralnik, E. (1995). *Goralnik Screening Test for Hebrew*. Even Yehuda: Matan.
- Gutiérrez-Clellen, V. F., Restrepo, M. A., and Simón-Cerejido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *J. Speech Lang. Hear. Res.* 49, 1209–1223. doi: 10.1044/1092-4388(2006/087)
- Håkansson, G., Salameh, E. K., and Nettelbladt, U. (2003). Measuring language development in bilingual children: Swedish–Arabic children with and without language impairment. *Linguistics* 41, 255–288. doi: 10.1515/ling.2003.009
- 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
- Hervé, C., and Serratrice, L. (2017). The development of determiners in the context of French–English bilingualism: a study of cross-linguistic influence. *J. Child Lang.* 45, 767–787. doi: 10.1017/S0305000917000459
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., and Parra, M. (2012). Dual language exposure and early bilingual development. *J. Child Lang.* 39, 1–27. doi: 10.1017/S0305000910000759
- Iluz-Cohen, P., and Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism* 16, 884–899. doi: 10.1017/S1366728912000788
- Janssen, B. (2016). *The Acquisition of Gender and Case in Polish and Russian: A Study of Monolingual and Bilingual Children*. Pegasus Oost-Europese Studies, 27. Amsterdam: Uitgeverij Pegasus.
- Janssen, B., Meir, N., Baker, A., and Armon-Lotem, S. (2015). “On-line comprehension of Russian case cues in monolingual Russian and bilingual Russian–Dutch and Russian–Hebrew children,” in *Proceedings of the 39th Annual Boston University Conference on Language Development, BUCLD 39*, eds E. Grillo and K. Jepson (Somerville, MA: Cascadia Press), 266–278.
- Jisa, H. (2000). Language mixing in the weak language: evidence from two children. *J. Pragmat.* 32, 1363–1386. doi: 10.1016/S0378-2166(99)00104-6
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., and Hulme, C. (2015). Sentence repetition is a measure of children’s language skills rather than working memory limitations. *Dev. Sci.* 18, 146–154. doi: 10.1111/desc.12202
- Kohnert, K. (2010). Bilingual children with primary language impairment: issues, evidence and implications for clinical actions. *J. Commun. Disord.* 43, 456–473. doi: 10.1016/j.jcomdis.2010.02.002
- 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
- Leonard, L. B. (2014). *Children with Specific Language Impairment*. Cambridge, MA: MIT press.
- Lust, B., Flynn, S., Blume, M., Park, S. W., Kang, C., Yang, S., et al. (2016). Assessing child bilingualism: direct assessment of bilingual syntax amends caretaker report. *Int. J. Biling.* 20, 153–172. doi: 10.1177/1367006914547661
- 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.
- Meir, N. (2017). Effects of Specific Language Impairment (SLI) and bilingualism on verbal short-term memory. *Linguist. Approaches Biling.* 7, 301–330. doi: 10.1075/lab.15033.mei
- Meir, N., and Armon-Lotem, S. (2015). “Disentangling bilingualism from SLI in Heritage Russian: the impact of L2 properties and length of exposure to the L2,” in *Proceedings of the GALA 2013, Language Acquisition and Development*, eds C. Hamann and E. Ruigendijk (Newcastle upon Tyne: Cambridge Scholars Publishing), 299–314.
- Meir, N., Walters, J., and Armon-Lotem, S. (2016). Disentangling SLI and bilingualism using sentence repetition tasks: the impact of L1 and L2 properties. *Int. J. Biling.* 20, 421–452. doi: 10.1177/1367006915609240
- Meir, N., Walters, J., and Armon-Lotem, S. (2017). Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguist. Approaches Biling.* 7, 514–553. doi: 10.1075/lab.15007.mei
- Meisel, J. M. (2007). The weaker language in early child bilingualism: acquiring a first language as a second language? *Appl. Psycholinguist.* 28, 495–514. doi: 10.1017/S0142716407070270
- Montrul, S. A. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*, Vol. 39. Amsterdam: John Benjamins Publishing. doi: 10.1075/sibil.39
- Müller, N., and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism* 4, 1–21. doi: 10.1017/S1366728901000116
- Müller, N., and Kupisch, T. (2003). Zum simultanen Erwerb des Deutschen und des Französischen bei (un) ausgeglichen bilingualen Kindern. *Vox Romanica* 62, 145–169.
- Novogrodsky, R., and Friedmann, N. (2006). The production of relative clauses in syntactic SLI: a window to the nature of the impairment. *Adv. Speech Lang. Pathol.* 8, 364–375. doi: 10.1080/14417040600919496
- Paradis, J., Crago, M., Genesee, F., and Rice, M. (2003). French-English bilingual children with SLI: How do they compare with their monolingual peers? *J. Speech Lang. Hear. Res.* 46, 113–127. doi: 10.1044/1092-4388(2003/009)
- 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
- Paradis, J., Rice, M. L., Crago, M., and Marquis, J. (2008). The acquisition of tense in English: distinguishing child second language from first language and specific language impairment. *Appl. Psycholinguist.* 29, 689–722. doi: 10.1017/S0142716408080296
- Peeters-Podgaevskaja, A. (2008). “Problemy osvoeniya russkogo yazyka kak vtorogo rodnogo det’mi 5–7 let i sozdanie adekvatnogo učebnogo posobija,” in *Literature and Beyond: Festschrift for Willem*, eds W. H. E. de Haard and G. Weststeijn (Amsterdam: Pegasus Oost-Europese Studies), 609–627.
- 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. (2018). *Heritage languages and Their Speakers*. Cambridge: Cambridge University Press.
- Raven, J. (1998). *Ravens Coloured Progressive Matrices*. San Antonio, TX: Psychological Corporation.
- Rice, M. L., Wexler, K., and Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *J. Speech Lang. Hear. Res.* 38, 850–863. doi: 10.1044/jshr.3804.850
- Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *Int. J. Lang. Commun. Disord.* 47, 499–510. doi: 10.1111/j.1460-6984.2012.00158.x
- Ringblom, N. (2012). *The Acquisition of Russian in a Language Contact Situation: A Case Study of a Bilingual Child in Sweden*. Doctoral dissertation, Acta Universitatis Stockholmiensis, Stockholm.
- Rodina, Y., and Westergaard, M. (2017). Grammatical gender in bilingual Norwegian–Russian acquisition: the role of input and transparency. *Bilingualism* 20, 197–214. doi: 10.1017/S1366728915000668
- Rothweiler, M., Schönenberger, M., and Sterner, F. (2017). Subject-verb agreement in German in bilingual children with and without SLI. *Z. Sprachwiss.* 36, 79–106. doi: 10.3109/02699206.2014.886726
- Schlyter, S. (1994). Early morphology in Swedish as the weaker language in French-Swedish bilingual children. *Scand. Work. Pap. Biling.* 9, 67–86.
- Schwartz, M., and Minkov, M. (2014). Russian case system acquisition among russian–hebrew speaking children. *J. Slav. Linguist.* 22, 51–92.
- Sorace, A., and Serratrice, L. (2009). Internal and external interfaces in bilingual language development: beyond structural overlap. *Int. J. Biling.* 13, 195–210. doi: 10.1177/1367006909339810
- Stokes, S. F., Wong, A. M., Fletcher, P., and Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: the case of Cantonese. *J. Speech Lang. Hear. Res.* 49, 219–236. doi: 10.1044/1092-4388(2006/019)
- Theodorou, E., Kambanaros, M., and Grohmann, K. K. (2017). Sentence repetition as a tool for screening morphosyntactic abilities of bilingual children with SLI. *Front. Psychol.* 8:2104. doi: 10.3389/fpsyg.2017.02104
- Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 15, 426–445. doi: 10.1177/1367006911403202
- 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. doi: 10.21832/9781783093137-014
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., and O’Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40, 1245–1260. doi: 10.1044/jslhr.4006.1245
- Turian, D., and Altenberg, E. P. (1991). “Compensatory strategies of child first language attrition,” in *First Language* eds H. W. Seliger and R. M. Vago (Cambridge: Cambridge University Press), 207–226.
- Unsworth, S. (2016). “Amount of exposure as a proxy for dominance in bilingual language acquisition,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalan and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173. doi: 10.1017/CBO9781107375345.008
- Yip, V., and Matthews, S. (2000). Syntactic transfer in a cantonese-english bilingual child. *Biling. Lang. Cogn.* 3, 193–208.

**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.

The reviewer LI and handling Editor declared their shared affiliation.

Copyright © 2018 Meir. 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.

## APPENDIX A: Structures Tested in the SRep Tasks Based on LITMUS-SRep Developed Within the COST Action IS0804 (Marinis and Armon-Lotem, 2015).

**TABLE A1** | Structures tested in the Russian SRep task with examples (number of sentences per structure in brackets).

Level	Structure	Example							
1	SVO (8)	<i>mal'čiki</i>	<i>myli</i>	<i>posudu</i>	<i>posle obeda.</i>				
		boys.NOM	washed.PL	dishes.ACC	after lunch				
		'A/The boys did dishes after lunch'							
	SOV (4)	<i>mal'čik</i>	<i>devočku</i>	<i>požalel</i>	<i>na uroke</i>				
		boy.NOM	girl.ACC	pitied.MASC.SG	on lesson				
		'A/The boy pitied a/the girl during the lesson'.							
OVS (4)	<i>kota</i>	<i>uvidela</i>	<i>myška</i>	<i>vo dvore.</i>					
	cat.ACC	saw.FEM.SG	mouse.NOM	in yard					
	'A/The mouse saw a/the cat in the yard'.								
2	Biclausal sentences with coordination (4)	<i>mama</i>	<i>ispekla</i>	<i>tort,</i>	<i>i</i>	<i>my</i>	<i>pozvali</i>	<i>druzej.</i>	
		mother.NOM	baked. FEM.SG	cake.ACC	and	we.NOM	called.PL	friends.ACC	
		'The mother made a cake and we invited friends'.							
	Biclausal sentences with subordination (4)	<i>devočka</i>	<i>vzjala</i>	<i>zontik,</i>	<i>potomu što</i>	<i>šjol</i>	<i>dožd'.</i>		
		girl.NOM	took.FEM.SG	umbrella	because	went.MASC.SG	rain.NOM		
		'A/They girl took an umbrella because it was raining'							
Object questions (4)	<i>kakuju</i>	<i>obez'janu</i>	<i>našjol</i>	<i>krokodil?</i>					
	which.FEM.ACC	monkey.FEM.ACC	found.MASC.SG	crocodile.MASC.NOM?					
	'Which monkey did a/the crocodile find?'								
Oblique questions (4)	<i>u</i>	<i>kakogo</i>	<i>mal'čika</i>	<i>sestra</i>	<i>vzjala</i>	<i>kraski?</i>			
	at	which.MASC.GEN	boy.GEN	sister.FEM.NOM	took.FEM.SG	paints.ACC			
	'From which boy did the sister take paints?'								
3	Real conditionals (4)	<i>my poedem</i>	<i>na</i>	<i>more,</i>	<i>esli</i>	<i>prosnemsja</i>	<i>rano.</i>		
		we go.1P.PL	on	sea,	if	wake-up.1P.PL	early.		
		'We will go to the sea, if (we) wake up early'.							
	Unreal conditionals (4)	<i>esli by</i>	<i>sestra</i>	<i>pobedila,</i>	<i>papa</i>	<i>kupil</i>	<i>by</i>	<i>cvety.</i>	
		if SUBJ	sister.NOM	won.FEM.SG	father	bought.MASC.SG	SUBJ	flowers.ACC	
		'We the sister had won, the father would have bought flowers'.							
Subject relatives (8)	<i>jeto lošadka,</i>	<i>kotoraja</i>	<i>dognala</i>	<i>korovu.</i>					
	this horse.FEM.NOM	which.FEM.NOM	caught-up.FEM.SG	cow.FEM.ACC					
	'This is a/the horse that caught-up-with a/the cow'.								
Object relatives (8)	<i>jeto</i>	<i>medved',</i>	<i>kotorogo</i>	<i>obmanula</i>	<i>lisa.</i>				
	this	bear.MASC.NOM	which.MASC.ACC	fooled.FEM.SG	fox.FEM.NOM				
	'This is a/the bear a/the fox fooled'.								

**TABLE A2 |** Structures tested in the Hebrew SRep task with examples (number of sentences per structure in brackets).

Level	Structure	Example			
1	SVO (8)	<i>hu axal/MASC.SG</i> he ate	<i>tapuax axarey</i> apple after	<i>ha-</i> DEF-	<i>ši'ur</i> lesson
	Bicausal sentences with coordination (4)	'He ate an apple after the lesson'. <i>ha-banot</i> DEF-girls	<i>qonot</i> buy.FEM.PL	<i>matanot ve</i> presents and	<i>ha-banim</i> DEF- boys
	Bicausal sentences with subordination (4)	'The girls buy presents and the boys play with a ball'. <i>ha- talmidot</i> DEF- pupils.FEM.PL	<i>ohavot</i> love.FEM.PL	<i>še-</i> that-	<i>ha- mora</i> DEF- teacher
2	Object questions (4)	'The pupils love that the teacher gets late'. <i>eyze</i> which.MASC	<i>garav</i> thief	<i>ha- shoteret</i> DEF- policewoman.FEM	<i>tafsa</i> caught.3P. FEM.SG
	Oblique questions (4)	'Which boy did the teacher frighten?' <i>im</i> with	<i>eyze</i> which.MASC	<i>xaver</i> friend.MASC	<i>ha- yeled</i> DEF- boy
	Object relatives (4)	'With which friend did the boy quarrel?' <i>zot</i> that	<i>ha-</i> DEF-	<i>yalda</i> girl.FEM.SG	<i>še</i> that
		'I saw the dog that the horse pushed'. <i>ra' inu</i> saw.1P.PL	<i>et</i> ACC	<i>ha-yeled</i> DEF- boy.MASC.SG	<i>še ha- sus</i> that DEF- horse
3	Real conditionals (4)	'(We) saw the boy that the horse kicked (into him)'. <i>im</i> if	<i>yifye</i> will+be	<i>xam nelex</i> hot will-go.1P.PL	<i>lišxot</i> to+swim
	Unreal conditionals (4)	'If it is hot, we will go to swim to the sea'. <i>ha- shaxen</i> DEF- neighbor.MASC.SG	<i>haya</i> was. MASC.SG	<i>qosher</i> tie.MASC.SG	<i>et ha-kelev</i> ACC DEF- dog
	Bicausal sentences with phrasal conjunctions (8)	'The neighbor would have tied up the dog if he were nice'. <i>kedey</i> in-order	<i>le'efot</i> to+bake	<i>uga</i> cake	<i>qaninu</i> bought.1P.PL
	VSO (8)	'In order to bake a cake, we bought milk and eggs'. <i>ba-</i> in+DEF	<i>qayic</i> summer	<i>halxu</i> went.3P.PL	<i>ha- yeladim</i> DEF- children.MASC.PL
		'In the morning the girl jumped on the rope'. <i>im hu haya nexmad.</i> if he was nice.		<i>xalav</i> milk	<i>im hu haya nexmad.</i> if he was nice.
				<i>u- beycim.</i> and eggs.	<i>u- beycim.</i> and eggs.
				<i>la- yam.</i> to+DEF- sea.	<i>la- yam.</i> to+DEF- sea.



# Bilinguals' Sensitivity to Grammatical Gender Cues in Russian: The Role of Cumulative Input, Proficiency, and Dominance

Natalia Mitrofanova<sup>1\*</sup>, Yulia Rodina<sup>1</sup>, Olga Urek<sup>1</sup> and Marit Westergaard<sup>1,2</sup>

<sup>1</sup> Department of Language and Culture, UiT The Arctic University of Norway, Tromsø, Norway, <sup>2</sup> Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway

## OPEN ACCESS

### Edited by:

Esther Rinke,  
Goethe-Universität Frankfurt am Main,  
Germany

### Reviewed by:

Zuzanna Fuchs,  
Harvard University, United States  
Elena Dieser,  
Universität Würzburg, Germany

### \*Correspondence:

Natalia Mitrofanova  
natalia.mitrofanova@uit.no

### Specialty section:

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

**Received:** 10 May 2018

**Accepted:** 14 September 2018

**Published:** 11 October 2018

### Citation:

Mitrofanova N, Rodina Y, Urek O and  
Westergaard M (2018) Bilinguals'  
Sensitivity to Grammatical Gender  
Cues in Russian: The Role  
of Cumulative Input, Proficiency,  
and Dominance.  
Front. Psychol. 9:1894.  
doi: 10.3389/fpsyg.2018.01894

This paper reports on an experimental study investigating the acquisition of grammatical gender in Russian by heritage speakers living in Norway. The participants are 54 Norwegian-Russian bilingual children (4;0–10;2) as well as 107 Russian monolingual controls (3;0–7;0). Previous research has shown that grammatical gender is problematic for bilingual speakers, especially in cases where gender assignment is opaque (Polinsky, 2008; Schwartz et al., 2015; Rodina and Westergaard, 2017). Furthermore, factors such as proficiency and family type (one or two Russian-speaking parents) have been argued to be important. Interestingly, previous findings differ with respect to the kind of errors children make: restructuring to a two-gender system (masculine–feminine, see Polinsky, 2008) or defaulting to masculine (see Rodina and Westergaard, 2017). It is also not clear to what extent children are sensitive to gender cues or whether certain agreement patterns are simply memorized. To investigate this, we used both existing nouns and nonce words and tested both transparent and opaque gender cues. The results were checked against a number of background factors measuring exposure, proficiency, and dominance. Our findings show that bilingual children are clearly sensitive to morphophonological cues for gender assignment. The most common and robust error pattern for all bilinguals involved overgeneralization to masculine (especially affecting neuter and opaque nouns). At the same time, children from families with two Russian-speaking parents and monolinguals also occasionally overused feminine with vowel-final nouns. The following variables were found to be the most reliable predictors of accuracy on grammatical gender tasks: cumulative length of exposure (CLOE) and consistency of input in Russian, as well as the presence of older siblings, with CLOE to Russian being by far the most robust and important predictor. Furthermore, we show that a lexical diversity measure (number of different words in a Russian narrative) is also correlated significantly with the children's performance on the gender tasks. At the same time, our results indicate that relative measures of dominance (e.g., the difference in exposure between the two languages or the difference in narrative scores) may be redundant when more robust absolute measures are present (CLOE and lexical diversity in the heritage language).

**Keywords:** nonce words, default gender, heritage speaker, Norwegian-Russian bilinguals, transparent/opaque gender, proficiency, cumulative length of exposure, lexical diversity



## INTRODUCTION

In this paper, we investigate heritage speakers' sensitivity to gender cues in Russian through a prism of a composite measure, combining linguistic background variables as well as measures of general proficiency and dominance. This novel method allows for a more direct way of measuring the predictive power of different variables for bilinguals' linguistic competence. The Russian three-gender system (masculine, feminine, and neuter) is relatively transparent, with some opaque cases, and it has been shown to be in place early in monolingual L1 acquisition. However, grammatical gender has been argued to be somewhat problematic for certain groups of heritage speakers, who have been found to develop a reduced gender system of only masculine and feminine (Polinsky, 2008) or no gender system at all, defaulting to masculine (Rodina and Westergaard, 2017). The factors that have been invoked to identify these groups of heritage speakers include general proficiency (Polinsky, 2008) as well as family type (one or two Russian-speaking parents) and amount of input (Rodina and Westergaard, 2017). In the current paper we use a much more detailed battery of 20 background variables as well as a proficiency measure based on semi-spontaneous narratives. In order to test whether heritage speakers are sensitive to morphophonological gender cues (and do not just memorize item-based patterns), we designed gender tasks that include both existing and nonce words. The participants for the study were 54 bilingual children growing up in Norway (age range 4;0–10;2) and 107 monolingual controls in Russia. The bilingual participants were from families with two Russian-speaking parents (the RR group) or families with one Russian- and one Norwegian-speaking parent (the NR group). The results show that, while there is considerable defaulting to masculine in the production of some of the heritage speakers, the general picture is that they are clearly sensitive to gender cues in the nonce word task. Furthermore, with respect to the background variables and proficiency measures, the statistical analysis shows that the best predictors of the children's performance on the gender tasks are a combination of three background variables (cumulative length of exposure (CLOE), consistency of input, and the presence of an older sibling) and one proficiency measure (lexical diversity in the narrative task). We argue that this shows that language dominance in heritage speakers is a relative concept that must take a number of factors into account in order to explain the acquisition of complex linguistic phenomena such as gender.

The paper is structured as follows: In the next section, we provide some background for the study, including a brief description of the gender system of Russian, an overview of previous research on the acquisition of gender in heritage language, and a discussion of commonly used proficiency and dominance variables. Section "Research Questions and Predictions" introduces our research questions and corresponding predictions based on previous findings, and Section "Materials and Methods" provides an overview of the participants of the study, the gender tasks, the background variables collected, as well as the language proficiency measures. In Section "Results," we present the results of the study and a detailed analysis in terms of a number of statistical models.

Section "Discussion" contains a discussion of our findings and Section "Conclusion" provides a brief conclusion.

## BACKGROUND

### Gender in Russian

Russian distinguishes between three grammatical genders – masculine, feminine, and neuter. Gender agreement is expressed as a suffix, and appears on singular adjectives, verbs in the past tense, demonstratives, participles, and certain pronouns. This is illustrated in (1). In the glosses, the gender of the noun is marked in parentheses and the agreeing item is marked after a full stop. In the present study, we only consider adjective-noun agreement in the nominative singular.

#### (1) Gender agreement marking in Russian

- |    |                     |                     |                         |
|----|---------------------|---------------------|-------------------------|
| a. | Moja                | bol'saja            | mašina                  |
|    | My. <sub>F</sub>    | large. <sub>F</sub> | car <sub>(F)</sub>      |
|    | 'My large car'      |                     |                         |
| b. | Moj                 | bol'soj             | slon                    |
|    | My. <sub>M</sub>    | large. <sub>M</sub> | elephant <sub>(M)</sub> |
|    | 'My large elephant' |                     |                         |
| c. | Mojo                | bol'soje            | krylo                   |
|    | My. <sub>N</sub>    | large. <sub>N</sub> | wing <sub>(N)</sub>     |
|    | 'My large wing'     |                     |                         |

The distribution of genders in the lexicon is uneven, with masculine nouns constituting approximately 46% of all nouns, feminines 41%, and neuters only about 13% (Corbett, 1991). Masculine is usually considered to be the default gender, since it is the most frequent, attracts most borrowings, and is associated with the default declension class (Corbett, 2007, p. 267). In addition, masculine agreement is used to refer to mixed-gender groups and in cases where the biological gender of an animate referent is unknown or unclear (Corbett, 2007, pp. 271–272).

Gender assignment in Russian is largely predictable, i.e., the grammatical gender of the noun is usually evident from its phonological shape in the nominative singular. Thus, nouns ending in non-palatal consonants are masculine (e.g., *stol* 'table'), nouns ending in stressed [a] are predominantly feminine (e.g., *noga* 'leg'), and nouns ending in stressed [o] are neuter (e.g., *steklo* 'glass'). Such nouns will be referred to as transparent. However, in certain cases the form of the noun in the nominative singular is opaque. For example, both feminine and masculine nouns may end in palatal and postalveolar consonants in the nominative singular (e.g., *gus'* 'goose.MASC,' *rys'* 'lynx.FEM'). Gender marking on nouns ending in palatalized consonants has been found to be problematic in monolingual first language acquisition, where overgeneralization to the masculine has been observed with feminine nouns during the preschool years (Gvozdev, 1961 based on diary data; Ceitlin, 2005, 2009 based on corpus data). This is likely due to the higher frequency of masculine nouns. It should be noted that the opposite, i.e., using feminine forms with masculine nouns ending in palatal consonants has not been attested in monolingual children. Other non-transparent nouns include

those ending in unstressed vowels. Due to the application of a vowel reduction process, underlying vowels /a/ and /o/ both get realized as [ə] in unstressed position, making nouns like *part[ə]* ('desk.FEM') and *sit[ə]* ('sieve.NEUT') opaque with respect to gender (see Iosad, 2012 on vowel reduction in Russian). Russian children have been shown to overgeneralize feminine agreement with non-transparent neuter nouns (Gvozdev, 1961; Popova, 1973). The opposite pattern, i.e., neuter agreement with stem-stressed feminines, has not been attested. All phonologically opaque nouns can be disambiguated by the case paradigm that they follow (e.g., *gus'-u* 'goose-MASC.DAT' vs. *rys'-i* 'lynx-FEM.DAT'). Thus, knowing the correlation between declensional class and gender is crucial in order to successfully predict the gender of these nouns.

Importantly, in monolingual acquisition, the masculine-feminine distinction is established very early, at approximately the age of 2 (Gvozdev, 1961; Ceitlin, 2005, 2009). Before their second birthday some children are reported to go through a short stage when feminine agreement is overgeneralized with masculine and neuter nouns (Gvozdev, 1961; Popova, 1973; Zakharova, 1973). Acquisition of neuter seems comparatively more difficult, which can be attributed to its low frequency in the input. While gender agreement with transparent neuters is usually mastered between 3;0 and 4;0 years of age, opaque neuters remain problematic until approximately the age of 6;0 (Gvozdev, 1961; Ceitlin, 2009).

The next section shows that gender marking has been found to be problematic for speakers of Russian as a heritage language. However, their overgeneralization patterns do not always match those of monolinguals.

## Gender Acquisition in Heritage Russian

Grammatical gender has been shown to be vulnerable in Russian heritage language, where both quantitative and qualitative differences have been observed in child and adult heritage speakers (e.g., Polinsky, 2008; Schwartz et al., 2015; Rodina and Westergaard, 2017). Non-target-like performance is mainly attributed to a combination of factors such as non-transparency of gender cues and insufficient exposure. To assess the role of participants' background, studies have typically employed different measures for children and adults. While the adults in Polinsky (2008) were assessed using a range of measures including a personal history questionnaire, a lexical translation task, and speech rate in oral narratives, in studies with children, family type and parental background questionnaires have been central (e.g., Gathercole and Thomas, 2005; Unsworth et al., 2014; Fhlannchadha and Hickey, 2017; Rodina and Westergaard, 2017). At the same time, specific domain knowledge is captured by custom-tailored experiments investigating gender marking with different subclasses of nouns.

Polinsky (2008) used a combination of production and comprehension tasks with Russian-speaking adults who assigned and judged gender marking on adjectives and possessive pronouns. The stimuli included 122 inanimate Russian nouns. Interestingly, language dominance and proficiency were introduced as different concepts in the study. All heritage speakers were defined as English-dominant simply based on

the fact that they lived in the United States and English was the language of the society. Yet, they had varying proficiency in Russian as measured by their speech rate in oral narratives and lexical access on a lexical translation task. The heritage speakers' performance on the gender tasks was found to correlate with their language proficiency. Heritage speakers with faster speech rates and lexical access, defined as high-proficiency speakers, had developed a target-like three-gender system of masculine, feminine, and neuter. In contrast, low-proficiency heritage speakers developed a reduced two-gender system of masculine and feminine, as they assimilated opaque as well as transparent neuter nouns to the feminine. Polinsky emphasizes that, while the observed restructuring was found to correlate with speech rate and lexical access, it did not correlate with a distinction proposed by Au and Romo (1997), whereby participants are divided into overhearers, intermediate, and more advanced speakers based on personal history questionnaires.

Studies investigating grammatical gender in child bilinguals are more numerous, with evidence obtained in different socio-cultural contexts. Schwartz et al. (2015) studied the development of gender agreement in 70 sequential bilinguals aged 4–5 acquiring Russian in the United States, Finland, Germany, and Israel. Based on parental reports, the children across these groups were argued to be Russian dominant at the age of testing, since they were born in families with two Russian immigrant parents and entered bilingual preschools around age 2–3. The knowledge of grammatical gender agreement between adjectives and head nouns was tested with the same elicitation procedure in all groups of bilinguals as well as younger (3- to 4-year-old) and older (4- to 5-year-old) monolinguals. The stimuli included 70 Russian nouns. The analysis of the children's errors did not reveal any qualitative differences between any of the bilingual groups and the monolinguals. However, the comparison of bilinguals with age-matched monolinguals revealed that the errors were more persistent in bilinguals, especially with feminine nouns ending in a palatalized consonant and with stem-stressed neuters. Thus, the acquisition of gender is delayed in these sequential bilinguals, even though they were classified as Russian-dominant. Schwartz et al. (2015) also suggest that the presence of the grammatical category gender in both languages of a bilingual facilitates acquisition, pointing out that the German-Russian and Hebrew-Russian bilinguals, whose majority language has grammatical gender, outperformed the English-Russian and Finnish-Russian bilinguals, whose majority language has no gender category.

Rodina and Westergaard (2017) investigated gender marking on adjectives in 20 simultaneous Norwegian-Russian bilinguals aged 4;1–7;11. The stimuli of the elicited production task included 30 Russian nouns. Bilingual family type was used as the main predictor variable in the study, since 10 children were from Russian-immigrant families and 10 children were from mixed Norwegian-Russian families. Importantly, the major difference in gender marking was found in a subset of five children from Norwegian-Russian families whose input in Russian was defined as very limited and inconsistent or mixed, since the children's Russian-speaking mothers reported using both languages and predominantly Norwegian with their children. In the gender

elicitation task, this subset of children used masculine agreement almost exclusively across all classes of nouns. The authors proposed that these children may be developing a variety of Russian with a more extensive reduction of the gender system (affecting both feminine and neuter, resulting in a system without gender), in contrast to the adults in Polinsky (2008), who showed signs of reduction of the neuter only. Like Polinsky (2008), Rodina and Westergaard (2017) suggested that this qualitative difference between monolinguals and heritage children could be due to the latter not having mastered the relatively complex declension system of Russian. These learners may thus be insensitive to the gender cues. The analysis of the bilingual data was based on two additional input measures – CLoE and the percentage of exposure at present (cf., the Bilingual Language Experience Calculator, Unsworth, 2013). Only CLoE was found to be a significant predictor of the bilingual children's gender marking in Russian, while the children's chronological age was the only significant predictor for gender accuracy in their majority language, Norwegian. This result was argued to support the conclusion that the amount of exposure was crucial for successful gender acquisition and that early exposure was not a sufficient condition.

Urek et al. (unpublished) used the procedure in Rodina and Westergaard (2017) to investigate gender acquisition in Latvian-Russian preschoolers resident in Riga, Latvia ( $N = 20$ , aged 4;0–6;10). In contrast to Rodina and Westergaard (2017), all the participants in this study come from mixed families, where one parent was a native speaker of the majority language and one a native speaker of the minority language. Crucially, the participants in this study reside in a country with a high degree of societal bilingualism and are therefore not heritage speakers of Russian *per se*. It was found that while the bilingual participants were less accurate in gender assignment than age-matched monolingual controls, they showed no evidence of restructuring or loss of the three-way gender contrast. However, just as in Rodina and Westergaard (2017), CLoE (controlling for age) was also found to be a significant predictor of accuracy.

## Assessment of Linguistic Proficiency, Input, and Dominance in Bilingual Acquisition

Bilingual speakers are a heterogeneous population, which is not surprising given that the input that children receive in the two languages can vary dramatically in terms of relative quantity, quality, and context (Sorace, 2005; De Houwer, 2007; De Cat and Serratrice, 2017). Apart from biographical variables such as the age of acquisition, chronological age, and place of birth, various measures have been proposed to quantify the amount of input that children receive, such as e.g., current amount of exposure (at home and at school etc.; cf., Gathercole and Thomas, 2009; Chondrogianni and Marinis, 2011), CLoE over time (Gutiérrez-Clellen and Kreiter, 2003; Blom, 2010; Unsworth, 2013), as well as richness and consistency of the input (Place and Hoff, 2011). Additional factors, such as the presence of siblings and birth order, language status (majority/minority) and language prestige (high/low),

daycare/school type (bilingual/monolingual/immersion), friends, literacy and literacy-related activities have also been shown to affect the linguistic development of bilingual children on a par with more general exposure variables (see Unsworth, 2013, 2015 and references therein). At the same time, several studies have highlighted correlations between the following so-called child-internal factors: the amount of output, MLU, vocabulary size, children's developing grammatical and phonological skills, fluency, and processing speed (Bohman et al., 2010; Paradis, 2011; Bedore et al., 2012).

Many of the aforementioned factors have been invoked in the discussion of dominance in bilinguals, and specifically of how dominance should best be measured. In many studies, the dominant language of a bilingual child is assumed to be the majority language of the wider community/country of residence (cf., Polinsky, 2008; see however, Schmeißer et al., 2015 for contrasting results). Alternatively, as argued by Unsworth (2015), current amount of exposure may be taken as a proxy for dominance/relative proficiency, while Treffers-Daller and Korybski (2015) propose that lexical diversity measures fit well as a means to operationalize dominance. Paradis et al. (2007) and Blom (2010) also take amount of input as the basis for determining the dominant language of a bilingual child, but also consider length of exposure since birth and amount of exposure in the home and at daycare/preschool/kindergarten. Bedore et al. (2012) apply a combination score of current language usage (current amount of exposure combined with children's own language output) as a proxy for dominance. Finally, Montrul (2015) argues for a more holistic, multidimensional approach to dominance, which includes all the three main components: biographical variables, proficiency, and input and use factors.

It should be noted that although language dominance and language proficiency are interrelated, they are nevertheless independent parameters. For example, while the dominance profiles may be similar in two groups of speakers, their absolute proficiency in the two languages may differ significantly (as is the case of e.g., Spanish L2 learners as compared to Spanish heritage speakers in the United States, see Montrul, 2015). Furthermore, as demonstrated by Schmeißer et al. (2015), high proficiency in a language does not imply that this language will necessarily be the dominant language for a bilingual child. Moreover, language dominance is not decisive when it comes to grammatical development, specifically cross-linguistic influence. As the authors argue, absolute rather than relative proficiency in the influenced language and the degree of complexity of the linguistic construction are much better predictors of cross-linguistic influence. Furthermore, contra what is commonly believed, the language of the country of residence does not always become the dominant language of a bilingual child, and the one-parent-one-language strategy is neither a necessary nor a sufficient prerequisite for balanced bilingualism. As the authors conclude, "more research on sociolinguistic factors, external to the child, which have been neglected in the past, is needed in order to help formulate recommendations for parents, doctors, and teachers, on how to promote high proficiency levels in the two languages of a bilingual" (Schmeißer et al., 2015, p. 64).

Following this line of research, the overarching goal of our current study is to investigate in detail the relative importance of the aforementioned factors for bilingual children's (rate of) grammatical acquisition in their minority language, specifically the acquisition of grammatical gender. Three groups of factors will be considered: language-internal factors (transparency of cues), child-external factors (e.g., current vs. cumulative exposure, relative difference in exposure between the two languages, parental language strategies, presence of siblings, etc.) as well as child-internal factors (children's performance skills on a narrative task, as well as the difference in their performance skills in the two languages).

## RESEARCH QUESTIONS AND PREDICTIONS

The present study examines bilingual Norwegian-Russian children's sensitivity to morphophonological gender cues in Russian, their minority language. In contrast to the previous studies reviewed in Section "Background," our experimental tasks employ both existing as well as novel nouns. This approach allows us to explore what mechanisms bilingual speakers use to assign gender and whether they develop a system of formal gender assignment rules. We also investigate the relationship between the bilinguals' knowledge of gender and background variables such as language exposure and language proficiency. The study addresses the following main research questions:

- (1) Do heritage speakers (HSs) of Russian differ significantly from monolinguals and in what conditions (on the real and nonce word tasks)?
- (2) Are there any differences between nouns with transparent and nouns with non-transparent morphophonological gender cues, for both monolinguals and HSs (on the real and nonce word tasks)?

Furthermore, by comparing the results of the real and nonce word tasks we aim to answer the following questions:

- (3) Do L1 children and HSs rely on their lexical knowledge of grammatical gender (i.e., is 'accuracy' on the real word task significantly higher or comparable with the accuracy on the nonce word task in transparent conditions, i.e., in cases where the phonological form of the noun straightforwardly predicts its gender)?
- (4) Is the reliance on lexically stored gender features stronger for HSs than for the L1 children (i.e., is there a significant interaction of task and group)?

One of the main purposes of our study is to consider in detail the background of the bilingual participants. We ask the following question:

- (5) Which background variables are the most reliable and robust in predicting children's performance on grammatical gender tasks?

In addition to the background variables, we assess the value of proficiency measures (narratives) in predicting HSs' performance on the gender assignment tasks. With respect to the contribution of the narrative proficiency measures we ask the following questions:

- (6) Do narrative proficiency variables such as lexical diversity correlate with HSs' performance on the gender assignment tasks?
- (7) Do these variables help to better predict the children's performance on the gender assignment tasks when used in combination with the background variables (i.e., is a model involving both proficiency and background variables statistically better at predicting the children's performance than a model involving only background variables)?

Finally, we ask whether dominance variables (operationalized as the difference in exposure to the two languages and the difference in proficiency on the narrative tasks between the two languages) can account for some part of the variance observed in the children's responses.

- (8) Do variables that quantify dominance help to better predict the children's performance on gender assignment tasks when used in combination with absolute exposure and proficiency variables (i.e., is a model involving both dominance and absolute exposure and proficiency variables statistically better at predicting the children's performance than a model involving only absolute exposure or proficiency variables)?

Based on the previous literature on bilingual language acquisition and the acquisition of gender in heritage Russian, we formulate the following predictions:

- (A) Based on the results of Polinsky (2008) we expect that bilingual children will significantly overgeneralize masculine agreement with nouns ending in a consonant and feminine agreement with nouns ending in a vowel.
- (B) Based on the results of Rodina and Westergaard (2017) we predict masculine overgeneralization to be the most pervasive error in the responses of bilingual children across all conditions. We also expect family type (two Russian-speaking parents vs. one Russian-speaking parent) to be a highly significant predictor of children's performance.
- (C) We expect that the amount of input will be a highly significant predictor of children's performance on the gender task. Following Paradis et al. (2007), Blom (2010), and Unsworth (2015), we expect CLoE and current exposure to be significant predictors of children's performance.
- (D) We expect that children's performance on the narrative tasks will correlate with their performance on the gender assignment tasks (cf., Montrul, 2015; Treffers-Daller and Korybski, 2015 for examples of correlations between lexical and grammatical abilities of bilinguals). We assume that the acquisition of nominal gender features is based



on the observation of nominal declension paradigms and agreement patterns, as well as generalizations over groups of nouns with shared morphological and phonological features. Thus, we predict that the children's lexical diversity scores will correlate positively with their performance on the real and nonce word tasks.

- (E) Finally, based on Schmeißer et al. (2015) we expect that absolute measures of children's exposure to Russian and their proficiency scores in Russian narratives will be better predictors of their performance on Russian gender assignment tasks than variables representing relative dominance (i.e., the difference in exposure to Russian and Norwegian and differences in the proficiency measures based on Russian and Norwegian narratives).

## MATERIALS AND METHODS

### Participants

For this study, we recruited 54 bilingual Russian-Norwegian children ( $N$  boys = 27) resident in Norway, ranging in age from 4;0 to 10;2 (mean age = 6;9). Of these, 22 children attend kindergarten, while the rest are schoolchildren. All participants in our study have a Russian-speaking mother and differ with respect to the first language of the father: 28 children (age range 4;3–9;9, mean age = 6;9) come from families with Norwegian-speaking fathers (and will be referred to as the NR group), while 26 children (age range 4;0–10;2, mean age = 6;9) come from families where both parents are Russian speakers (the RR group). All children included in this study were either born in Norway or arrived in Norway before the age of three. All come from middle-class households, where the education of the majority of the parents is at the level of an undergraduate degree. The bilingual participants were recruited and tested at Russian clubs in Oslo and Tromsø. These clubs offer weekly meetings for Russian-speaking children and provide classes on Russian language and culture (taught in Russian), as well as an informal socializing platform for Russian-speaking children and their families.

In addition, a group of monolingual controls ( $N$  = 107) ranging in age between 3;0 and 7;0 years (mean age = 5;2) were recruited and tested in Moscow and Ivanovo, Russia. All the monolingual children attended kindergarten at the time of testing.

### Gender Assignment Tasks

To examine bilingual Norwegian-Russian children's sensitivity to morphophonological gender cues in Russian we used two production tasks eliciting adjectival agreement with real nouns (Experiment 1) and nonce nouns (Experiment 2). The procedure used in both experiments was an adapted version of the picture-based elicitation task from Rodina and Westergaard (2013, 2017). The elicitation materials consisted of two sets of colored pictures. The pictures used in the real-word experiment (Experiment 1) were obtained from the Colourbox database; the pictures used in the nonce-word experiment (Experiment 2) were selected from the set of pictures of novel objects included in the Novel Object and Unusual Name Database (NOUN; Horst and Hout, 2016). The pictures used in Experiment 2 all depicted inanimate countable objects of variable shapes and textures.

The stimuli used in Experiment 1 consisted of 30 picturable nouns denoting everyday objects and animals assumed to be familiar to children at the relevant age. The nouns were evenly distributed across the three genders. In addition, the nouns of each gender varied with respect to morphophonological transparency, resulting in six conditions. Examples of the stimuli are given in **Table 1**.

The stimuli used in Experiment 2 were 25 novel nouns constructed to conform to Russian phonotactics. In order to avoid neighborhood density effects, only nouns that had no nominal phonological neighbors were selected. To achieve this, we used the Phonological Corpus Tools software (PCT; Hall et al., 2016) to check for any minimal pairs with the nouns included into the Frequency Dictionary of Russian (Sharoff, 2002). The novel nouns were equally distributed across five conditions, illustrated in **Table 2**. M-transparent, F-transparent, and N-transparent contained nouns with transparent masculine, feminine, and neuter cues respectively. The F/N-opaque condition contained stem-stressed vowel-final nouns (recall that in Russian these are ambiguous between feminine and neuter). The F/M-opaque condition contained nouns ending in palatal consonants (ambiguous between feminine and masculine).

In both experiments, two pictures of the same object differing in color were presented side by side on a laptop screen. The experimenter named the depicted object and then asked the participant to name the two objects along with their colors. The experimenter then pressed a button causing one of the pictures to disappear and asked the participant to identify the object that

**TABLE 1 |** Real noun stimuli.

Condition	M-transparent	F-transparent	N-transparent	M-opaque	F-opaque	N-opaque
Example	most 'bridge'	lisá 'fox'	vedró 'bucket'	gus' 'goose'	kost' 'bone'	myl[a] 'soap'

**TABLE 2 |** Novel noun stimuli.

Condition	M-transparent	F-transparent	N-transparent	F/N-opaque	F/M-opaque
Example	punip	kluvá	garpó	prúz[a]	dron'

disappeared. Thus, three instances of adjectival agreement were elicited for each target noun. Lead-in sentences were formulated in such a way as to avoid providing cues to the grammatical gender of the target noun. To familiarize the participants with the task, the test trials were preceded by two practice trials in both experiments. During the practice trials, plural forms were used to avoid priming. The elicitation procedure with a nonce stimulus noun is illustrated in (2):

(2) Elicitation procedure

- E.: Eto nazyvaetsja *punip*. Posmotri, čto zdes'?
- "This is called *punip*. Look, what is here?"
- C.: Eto krasnyj *punip*, a eto goluboj *punip*.  
this red.<sub>M</sub> *punip*<sub>(M)</sub>, and this blue.<sub>M</sub> *punip*<sub>(M)</sub>.  
"This is a red *punip*, and this is a blue *punip*."
- E: Čto sejčas propalo?
- "What has disappeared now?"
- C: Krasnyj *punip*.  
red.<sub>M</sub> *punip*<sub>(M)</sub>  
"The red *punip*."

All participants were tested individually by an experimenter who is a native speaker of Russian. The responses were audio-recorded and later transcribed and coded by the authors of this study.

## Background Variables

Background variables for the bilingual participants were obtained with the help of the Bilingual Language Experience Calculator (BiLEC, Unsworth, 2013), a parental questionnaire containing a set of questions designed to elicit detailed biographical data and information pertaining to the present language environment of a multilingual child in both languages, including exposure, context and use, as well as the child's linguistic experience from the onset of acquisition. BiLEC maps, *inter alia*, the proportion of input the child receives in each of the languages (both inside and outside of the home), the proportion of the child's own production in the L1 and the L2, and language exposure during holidays. It also includes questions on perceived receptive and productive language proficiency of the child and other members of the household (as reported by the respondent). BiLEC comes with an algorithm that automatically calculates numeric values for a range of pre-determined variables.

In the standard procedure, BiLEC serves as the basis for a parental interview. However, for the purpose of this study, BiLEC was translated into Russian and adapted into a questionnaire format in order to simplify data collection. The BiLEC questionnaires were filled out individually by one of each participant's parents (typically the mother). The responses were then entered into the BiLEC algorithm, and the values for a range of background variables were obtained.

The variables selected for the statistical analysis fall into three broad categories: biographic variables, language exposure, context and use variables, as well as maternal input variables. The biographic variables include age in months, family type (NR or RR), group (daycare or school), place of residence (Tromsø or Oslo), and the presence of siblings (younger and older).

The numeric values for the exposure variables were calculated automatically by the BiLEC algorithm (see Unsworth, 2013 for a detailed explanation of the calculations). Traditional length of exposure to Russian and Norwegian was calculated as the time elapsed from the date of first exposure to the date of testing. Thus, the traditional length of exposure to Russian corresponded to chronological age for all the children in our sample, while the length of exposure to Norwegian only corresponded to chronological age in children coming from NR families and varied for RR children (usually depending on when the child started attending daycare). Present weekly exposure to Russian/Norwegian was calculated as a proportion determined by dividing the total number of hours per week with exposure to Russian/Norwegian by the total number of waking hours each week. We included both 'present exposure at home' (only taking into account the proportion of Russian/Norwegian the child was exposed to in the household) and 'overall present exposure' (taking into account the overall weekly proportion of Russian/Norwegian the child was exposed to at home, school, and out-of-school activities including holidays). CLoE to Russian/Norwegian (in years) was calculated as the sum of proportions of each year in the child's life so far that included exposure to Russian/Norwegian. This measure takes into account how much each member of the household spoke each of the languages to the child during each year of the child's life so far, the amount of Russian/Norwegian spoken at the daycare/school the child attended, and the amount of Russian/Norwegian encountered during holidays.

In addition, three variables characterizing maternal language input were considered: consistency of input in Russian (binary variable indicating whether or not the mother reported using exclusively Russian when speaking to the child), proportion of Russian input from the mother (numeric variable estimated by the parent), and maternal productive proficiency in Norwegian (self-reported using a 6-point scale from 0 'do not speak at all' to 5 'native-like productive proficiency').

## Language Proficiency

Language proficiency was assessed in both Russian and Norwegian for a subset of bilingual children in our sample ( $N = 27$ ). We used the Multilingual Assessment Instrument for Narratives (MAIN, Gagarina et al., 2012) to elicit semi-spontaneous production samples. MAIN is a picture-based tool which contains four parallel stories ("Cat," "Dog," "Baby birds," and "Baby goats"), each illustrated with a six-picture sequence. MAIN was chosen for the present study since it is highly suitable for the elicitation of semi-spontaneous production samples in both of the languages of bilingual children, especially between the ages of 4 and 10.

We used the model story procedure to collect production samples in Norwegian and Russian (cf., Rodina, 2017). The child first heard a pre-recorded model story while looking at the picture sequence "Cat" or "Dog" and then answered 10 comprehension questions listed in the MAIN manual. This was done in order to establish contact with the child and to provide an example of narrative production. The child was then asked to narrate a

new story, either “Baby birds” or “Baby goats.” All the bilingual participants were tested in Russian first. Norwegian samples were collected approximately 2 weeks later. “Cat” and “Baby birds” scenarios were used to collect Russian narratives, and “Dog” and “Baby goats” scenarios were used for Norwegian narratives. The children were tested by research assistants who were native speakers of the respective languages. The children were tested individually, and their responses were audio-recorded and later orthographically transcribed.

In the analysis, we included two lexical measures of proficiency in each language sample: total number of words (i.e., all word tokens, TNW) and number of different words (i.e., word types, NDW). Mazes, repetitions, and incomplete utterances were excluded from the analysis. Both TNW and NDW have been shown to be important predictors of language development across different studies, including a previous investigation of narrative abilities in Norwegian-Russian bilingual preschoolers (Rodina, 2017).

## RESULTS

We start by presenting the results of Experiment 1 (real words, subsection “Experiment 1: Real Words”) and Experiment 2 (nonce words, subsection “Experiment 2: Nonce Words”). In subsection “Background Variables,” we summarize the effects of various background and proficiency variables on the children’s performance on the gender assignment tasks.

### Experiment 1: Real Words

**Figure 1** presents the accuracy in gender marking across the six experimental conditions (**Table 1**) and three participant

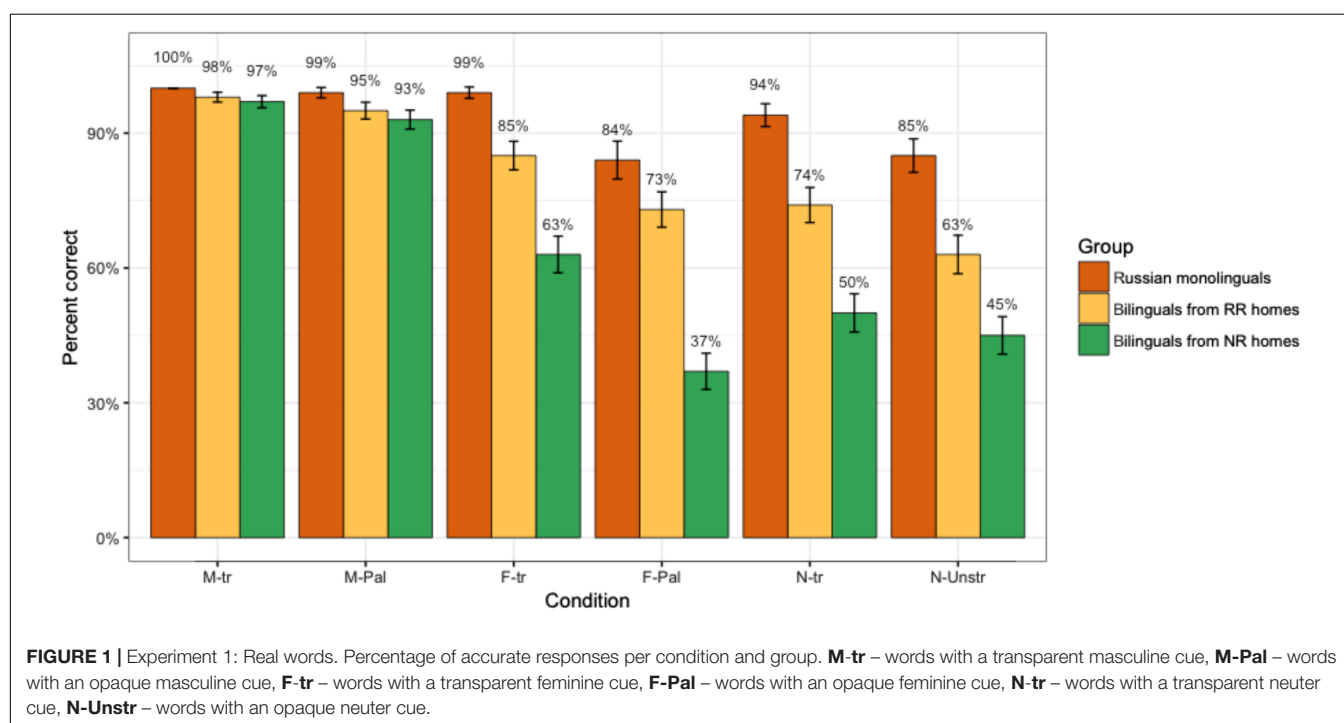
groups: Russian monolingual children, bilingual children from RR homes and bilingual children from NR homes. The accuracy rates of Russian monolinguals reveal that gender assignment is at-ceiling in M-transparent, F-transparent, and N-transparent as well as M-opaque conditions. Some non-target-like performance is observed in F-opaque and N-opaque conditions, where the accuracy rates are 85% and 86% respectively. Bilinguals from RR homes appear to be a close match to the monolinguals: F-opaque and N-opaque conditions are at 77% and 68% accuracy. However, some errors are found in the N-transparent condition, where the accuracy is 80%. Bilinguals from NR homes behave at-ceiling only in the M-transparent and M-opaque conditions. Their accuracy rates in all other conditions are below 60%.

We fit a generalized linear mixed logistic regression model where the binary variable accuracy was predicted by the interaction of Condition and Family (RR vs. NR vs. Monolingual R). Participants and items were included as random intercepts. To compare the groups within conditions, we conducted *post hoc* pairwise comparisons with the help of the R<sup>1</sup> package *lsmeans* (Lenth, 2016).

*Post hoc* pairwise comparisons of the groups within conditions revealed the following contrasts:

- (1) In the M-transparent and M-palatal conditions, there were no significant differences between the groups.
- (2) In the F-transparent and F-palatal conditions, the NR children were significantly less accurate than the RR children ( $p = 0.002$  and  $p < 0.001$ ,

<sup>1</sup> All models were fit using R version 3.4.4 (release 2018-03-15)



respectively) and monolinguals ( $p < 0.001$  in both conditions).

- (3) In the N-transparent condition, the NR group performed significantly less accurately than the RR children and monolingual controls ( $p < 0.01$  in both cases).
- (4) In the N-unstressed condition, the RR group patterned with the monolingual controls, while the NR group performed significantly less accurately (NR vs. Monolinguals:  $p = 0.01$ ; NR vs. RR:  $p = 0.03$ ).

*Post hoc* comparisons of different conditions within groups revealed the following contrasts:

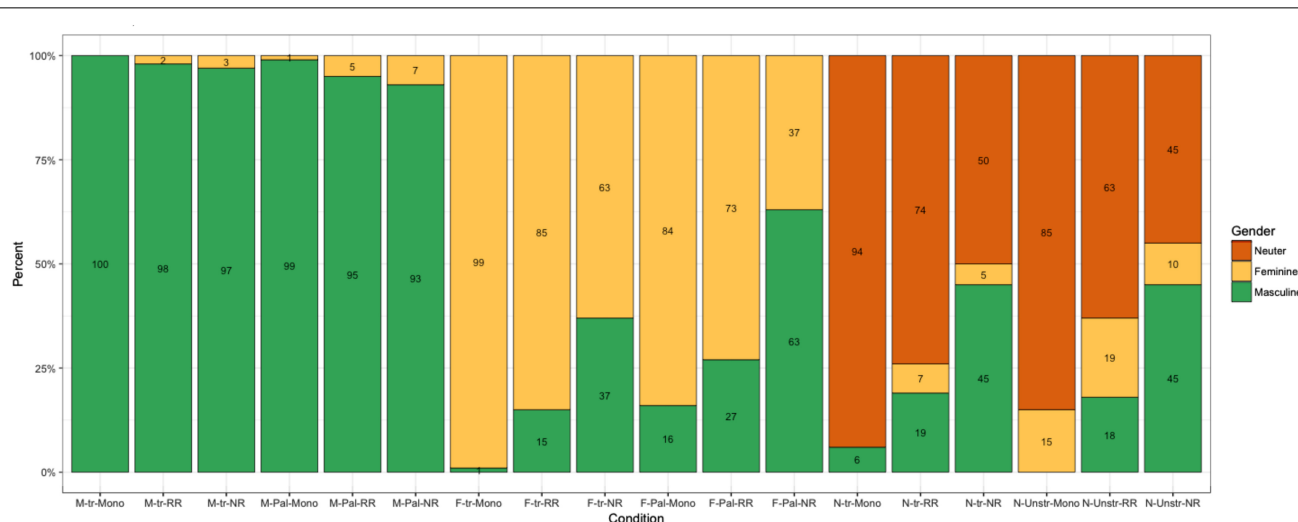
- (1) Monolingual children were (a) significantly more accurate on M-transparent and M-palatal than on N-unstressed ( $p < 0.001$  in both cases) and N-transparent conditions ( $p < 0.001$  in both cases); (b) significantly less accurate on the F-palatal than on the M-palatal condition ( $p = 0.01$ ) and marginally less accurate on the F-palatal than on the M-transparent condition ( $p = 0.08$ ); (c) significantly more accurate on the F-transparent condition than on the N-transparent and Neuter-unstressed conditions ( $p = 0.002$  and  $p < 0.001$ , respectively) and on the Feminine-palatal condition ( $p = 0.03$ ).
- (2) For the RR group, performance on the M-palatal condition was significantly better than on the F-palatal, N-transparent and N-unstressed conditions ( $p < 0.001$  in all cases) as well as the F-transparent condition ( $p = 0.03$ ). They were also significantly more accurate on the M-palatal condition than on the F-palatal, N-transparent and N-unstressed conditions ( $p < 0.001$  in all cases). Finally, they were significantly more target-like with

respect to F-transparent nouns than F-palatal nouns ( $p = 0.05$ ).

- (3) For the NR group, performance on M-transparent and M-palatal conditions was significantly more accurate than on all other conditions ( $p < 0.001$  in all cases). Accuracy on the F-transparent condition was significantly higher than on the F-palatal ( $p < 0.001$ ) and N-unstressed ( $p = 0.002$ ) conditions.
- (4) No other differences were significant.

**Figure 2** illustrates the use of masculine, feminine, and neuter agreement across all conditions and participant groups. The most common overgeneralization pattern in bilinguals involves the overuse of masculine agreement in all non-masculine conditions (F-opaque, F-transparent, N-opaque, N-transparent). This pattern is significantly more pronounced in the NR group than in the RR group. The NR group resorts to masculine across all non-masculine conditions (between 42% and 65% of the time), while the RR group overuses masculine significantly less (between 11% and 23% of the time) across all non-masculine conditions. Monolinguals erroneously use masculine 11% of the time, and only in the F-opaque condition, which bears an ambiguous feminine/masculine cue (final palatal consonant). In the N-opaque condition, where the phonological cue on the noun is ambiguous between feminine and neuter (final unstressed vowel), monolinguals overuse feminine (14% of the time), NR resort to masculine (in 51% of their responses), while RR children show both patterns (use feminine in 12% and masculine in 25% of the cases).

To sum up, in the real word experiment, we observe that the NR bilinguals are significantly different from Russian monolinguals and RR bilinguals. For all participant groups,



**FIGURE 2 |** Experiment 1: Real words. The use of masculine, feminine and neuter agreement per condition (in %): **M-tr** – words with a transparent masculine cue, **M-Pal** – words with an opaque masculine cue, **F-tr** – words with a transparent feminine cue, **F-Pal** – words with an opaque feminine cue, **N-tr** – words with a transparent neuter cue, **N-Unstr** – words with an opaque neuter cue. **Mono** – monolingual Russian children, **RR** – bilingual Norwegian-Russian children from families with two Russian-speaking parents, **NR** – bilingual Norwegian-Russian children from families with one Russian-speaking parent.



the M-transparent and M-opaque conditions are unproblematic, while the F-opaque and the N-opaque conditions pose the most difficulty.

## Experiment 2: Nonce Words

Recall from Section “Gender Assignment Tasks” that the nonce word experiment had five experimental conditions. In the three transparent conditions (M, F, N) we expected the use of masculine, feminine, and neuter agreement. In the opaque condition, two agreement options were possible: masculine and feminine in the FM condition and feminine and neuter in the FN condition.

We first present the results for the M-, F-, and N-transparent conditions in **Figure 3**, which compares the performance of all participant groups across these conditions in the nonce and real word tasks. **Figure 3** shows that in the three transparent conditions, children from all groups assign gender more ‘accurately’ (i.e., in accordance with the respective morpho-phonological cues) to real words than to nonce words. A generalized linear mixed effects regression analysis reveals that the ‘accuracy’ with feminine and neuter nouns is significantly higher in the real word task than in the nonce word task for all three groups of participants. Children use more masculine agreement in non-masculine conditions in the nonce-word task than in the real-word task. No significant interaction of Task and Group was found.

**Figure 4** illustrates the use of agreement in the nonce word experiment in all conditions. As **Figure 4** shows, the most common overgeneralization pattern observed in the bilingual groups is the overuse of masculine in all non-masculine conditions (similarly to the real word task). Notice also that the N-transparent condition turned out to be quite problematic for the NR and RR groups. Children from these two groups

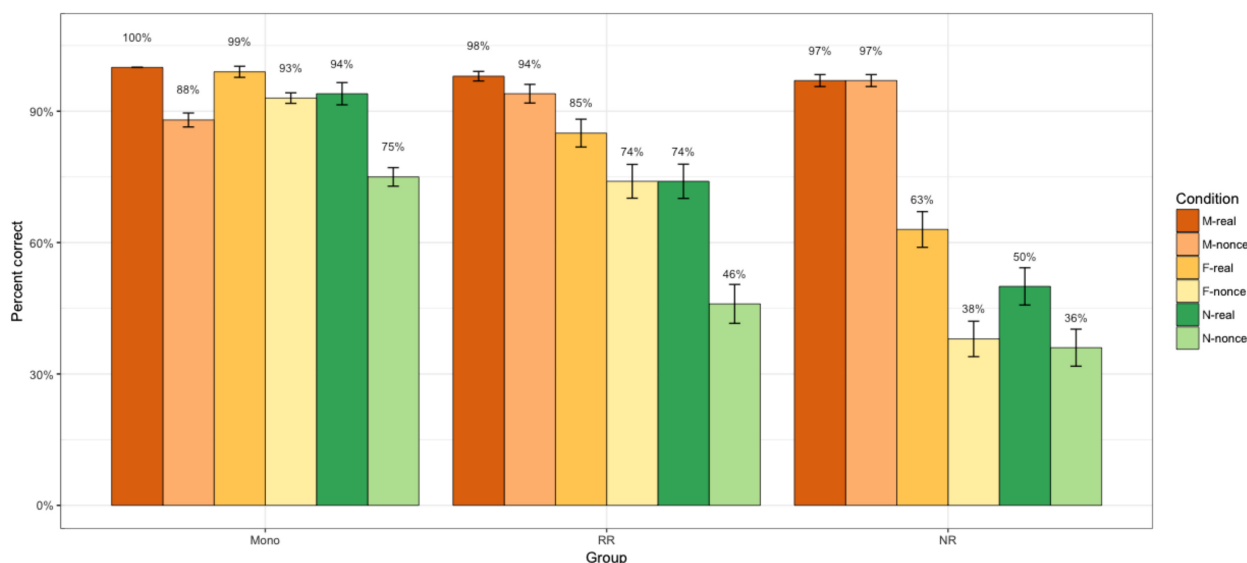
produced neuter agreement in 32% and 48% of the cases, respectively, while monolinguals assigned neuter in 75% of the cases.

To analyze the differences between the groups, and more specifically, between NR and RR children in comparison with the monolingual controls, we fit a generalized linear mixed logistic regression model to predict the probability of using masculine agreement by the interaction of Condition and Family. Participants and Items were included as random intercepts.

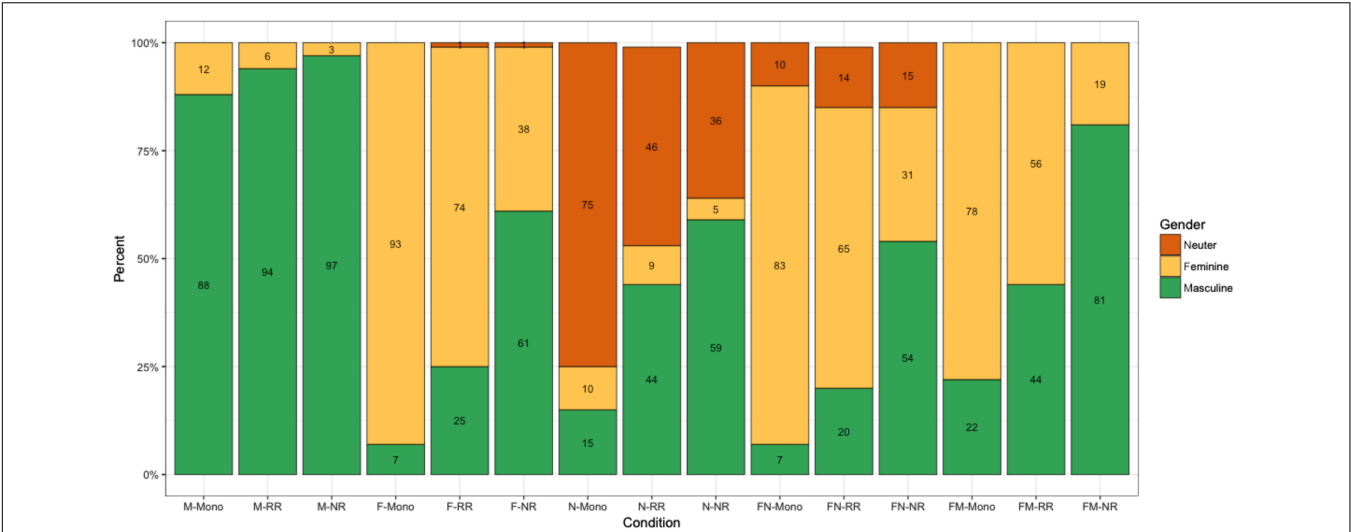
*Post hoc* pairwise comparisons of the performance of the groups within conditions revealed the following contrasts.

- (1) In the M-transparent condition, the NR group differed significantly from the Russian monolingual group ( $p = 0.003$ ).<sup>2</sup>
- (2) In the F-transparent condition, all groups differ significantly from each other, with the most significant contrasts being between the NR group and the monolinguals (RR vs. Monolinguals:  $p = 0.02$ , NR vs. Monolinguals:  $p < 0.0001$ , NR vs. RR:  $p < 0.001$ ).
- (3) In the N-transparent condition, the RR and NR bilingual groups differed significantly from the monolingual controls ( $p = 0.007$  and  $p < 0.0001$ , respectively).
- (4) In the two opaque conditions (F/N and F/M conditions), the RR group patterned with the monolingual controls, while the NR group used significantly more masculine than the two other groups ( $p < 0.001$  for all contrasts).

<sup>2</sup>This was a rather unexpected result, which can be attributed to the fact that a small group of 3-year-old monolinguals defaulted to feminine across the board in this task, thus lowering the overall proportion of M in this condition, while defaulting to F was not observed in the bilingual groups.



**FIGURE 3 |** Experiments 1 and 2: Gender assignment in accordance with transparent masculine, feminine, and neuter morphonological cues on real and nonce word tasks.



**FIGURE 4 |** Experiment 2: Nonce words. The use of masculine, feminine and neuter agreement per condition (in %): **M** – words with a transparent masculine cue, **F** – words with a transparent feminine cue, **N** – words with a transparent neuter cue, **FN** – words with an opaque feminine/neuter cue, **FM** – words with an opaque masculine/feminine cue. **Mono** – monolingual Russian children; **RR** – bilingual Norwegian-Russian children from families with two Russian-speaking parents; **NR** – bilingual Norwegian-Russian children from families with one Russian-speaking parent.

- (5) With respect to the agreement patterns, RR children performed similarly to monolinguals in that they preferred feminine in these conditions (83% and 69% feminine in monolingual and RR groups in the F/N condition, and 78% and 61% feminine in the F/M condition, respectively). On the other hand, NR children preferred masculine in both F/N and F/M conditions (57% and 83% masculine, respectively), while feminine was not the most frequent choice (28% and 17%, respectively).
- (6) No other differences were significant.

Background Variables

One of the goals of our study was to estimate which of the background variables were the most robust and reliable predictors of the children’s performance on Russian gender assignment tasks. To do so, we applied a non-parametric approach (random forests analysis), in combination with standard generalized mixed effects linear regression modeling. We included 20 independent variables calculated with the help of BiLEC (Unsworth, 2013), which we had collected with the parents of the 54 bilingual participants (abbreviations used in the analysis and in Figures 5–8 below are provided in the rightmost column):

i.	Traditional length of exposure to Russian	Trad_LoE_RUS
ii.	Traditional length of exposure to Norwegian	Trad_LoE_NOR
iii.	Cumulative length of exposure to Russian	Cum_LoE_RUS

iv.	Cumulative length of exposure to Norwegian	Cum_LoE_NOR
v.	Weekly exposure to Russian at home (at present)	Exp_week_home_RUS
vi.	Weekly exposure to Norwegian at home (at present)	Exp_week_home_NOR
vii.	Weekly exposure to Russian at home, school and out-of-school activities (at present)	Exp_week_hse_RUS
viii.	Weekly exposure to Norwegian at home, school and out-of-school activities at present	Exp_week_hse_NOR
ix.	Consistency of input in Russian (Yes, if the mother <sup>3</sup> indicated that she used Russian always or almost always with the child, No in all other cases)	RUS_consistent
x.	Proportion of Russian with mother	prop_RU_mother
xi.	Maternal proficiency in Norwegian	mother_NO_speaking
xii.	Age in months	Age_months
xiii.	Family type (NR vs. RR)	Family
xiv.	Group (daycare/school)	Group
xv.	Presence of older siblings	older_sibling
xvi.	Presence of younger siblings	younger_sibling
xvii.	Exposure to Russian during holidays (calculated as weekly exposure to Russian at home, school, extra and holidays at present minus weekly exposure to Russian at home, school and extra)	Exp_week_hd_RUS

<sup>3</sup>All NR children in our sample had a Russian-speaking mother and a Norwegian-speaking father. The fathers reported no or low proficiency in Russian. In the analysis, we included three “maternal” variables: consistency and proportion of Russian input (with mother), as well as maternal proficiency in Norwegian (as it might be expected that that the more fluent the mother is in Norwegian, the more likely she would be to use Norwegian at home with other family members and her children).

xviii.	Exposure to Norwegian during holidays (calculated as weekly exposure to Norwegian at home, school, extra and holidays at present minus weekly exposure to Norwegian at home, school and extra)	Exp_week_hd_NO
xix.	Differences in current amount of exposure to Norwegian and Russian (calculated as Exposure to Russian at home, school and extra per week subtracted from Exposure to Norwegian at home, school and extra per week)	Diff_Exp_hse
xx.	Differences in cumulative amount of exposure to Norwegian and Russian (calculated as CLoE to Russian subtracted from CLoE to Norwegian)	DiffCumLoE

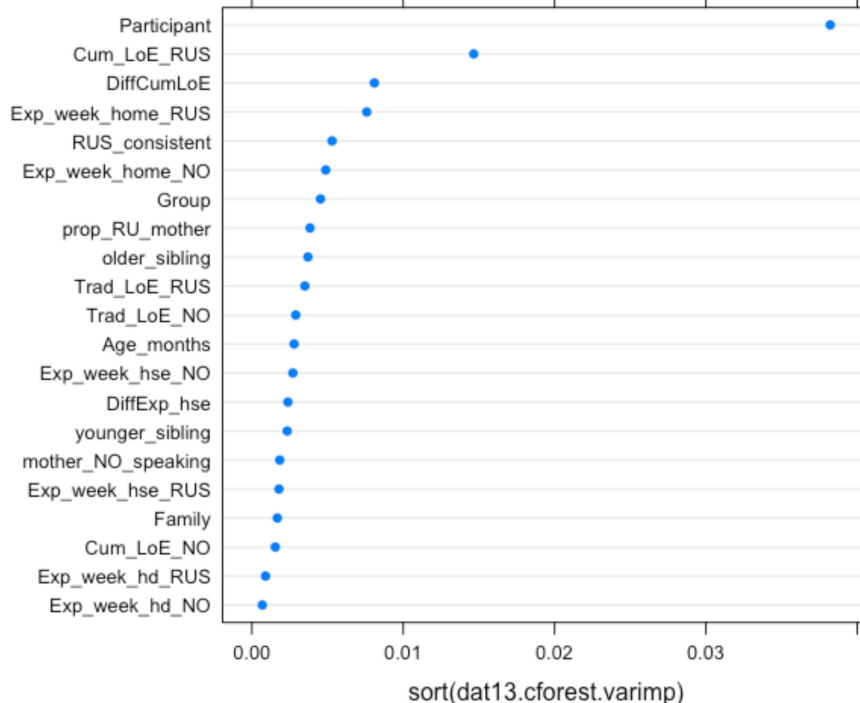
See **Table 3** for the descriptive statistics of the background variables.

To assess the effect of the children's background on their performance on the gender assignment tasks, we chose two binary dependent variables: accuracy and the probability of using masculine agreement in non-masculine conditions (masculine default). Note that in the opaque conditions of the nonce word experiment we coded both F and N responses in the FN condition and both F and M responses in the FM condition as 'accurate.'

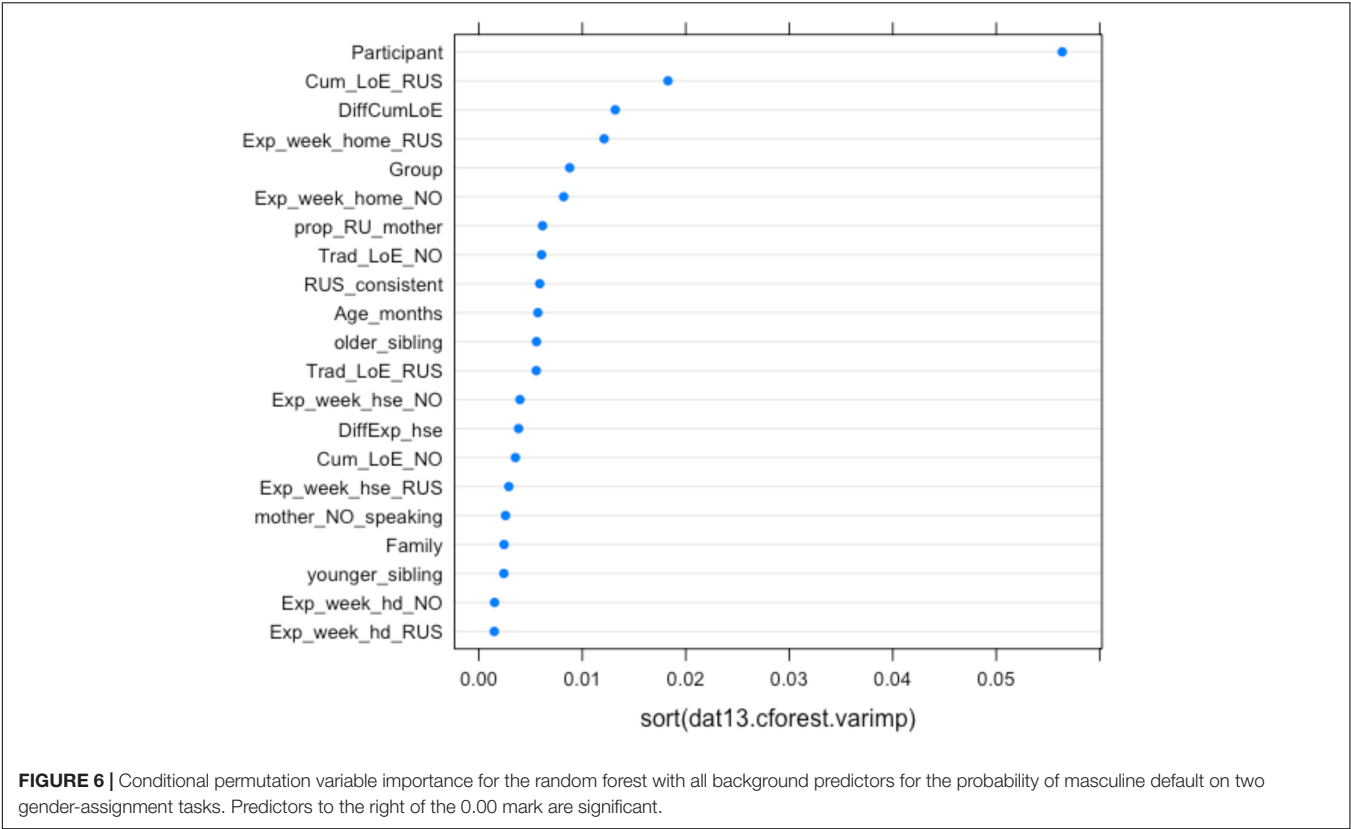
A serious challenge with data like ours has to do with the presence of many overlapping background variables. For example, exposure to Russian/Norwegian at home is collinear with exposure to Russian/Norwegian at home, school, and out-of-school activities; the amount of the child's exposure to Russian at home is negatively correlated with their amount of exposure to Norwegian at home; Family type (NR vs. RR home) has a direct impact on the amount of input in Russian and Norwegian that the child receives at home; proportion of Russian with the mother inevitably correlates with other variables concerning input in Russian, etc.

One possible way to cope with multiple collinear predictors is to apply dimension reduction techniques, such as principal components analysis, and then use standard regression with the reduced set of variables (see e.g., Strobl et al., 2009). However, principal components analysis would only be appropriate for numeric variables and cannot be applied to variables of other types, e.g., factors, in our case: the presence/absence of older/younger siblings, consistent/inconsistent input in Russian etc. Furthermore, as argued in Strobl et al. (2009, p. 324), dimension reduction techniques have "the disadvantage that the original input variables are projected onto a reduced set of components, so that their individual effect is no longer identifiable."

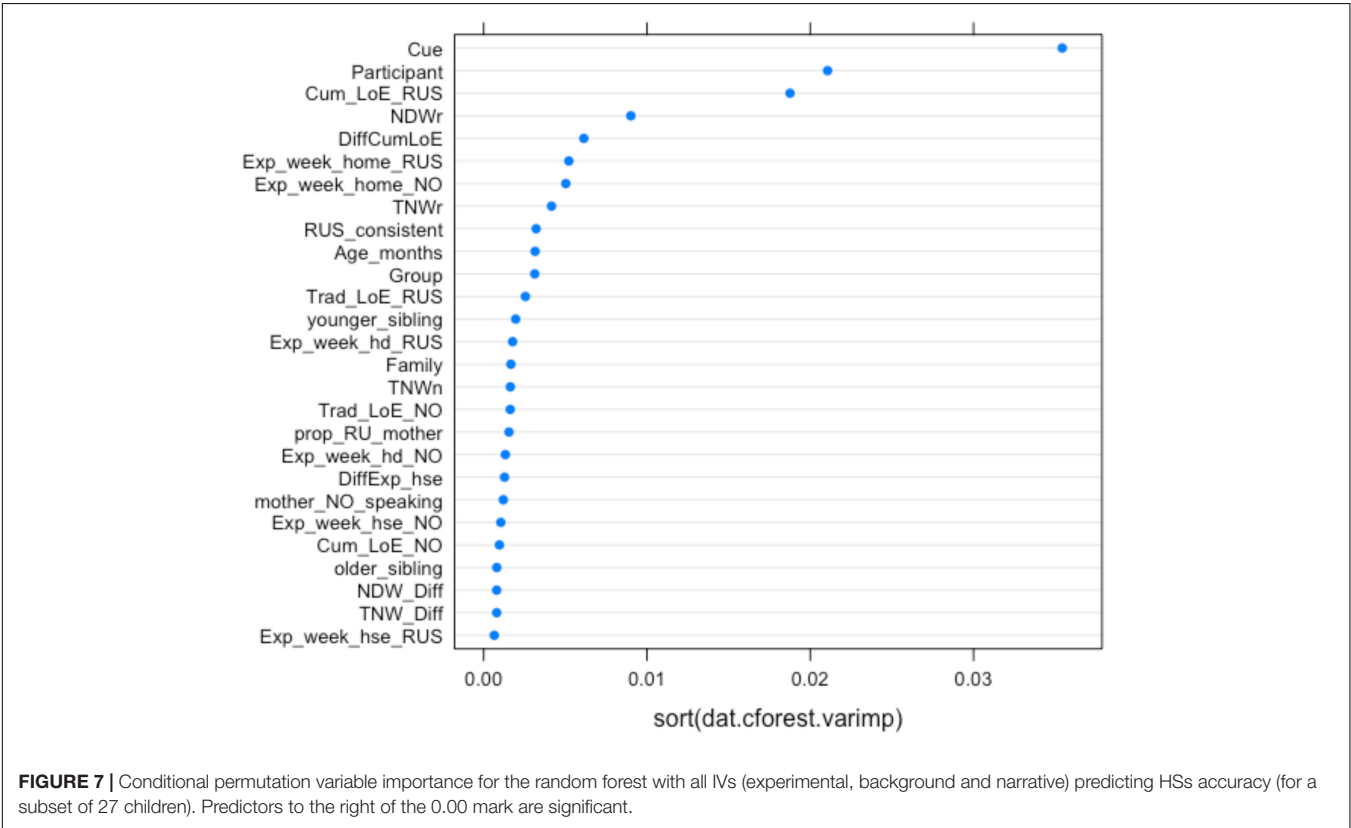
To overcome these limitations, we first ran a random forests analysis to estimate the relative importance of the different variables (see Breiman et al., 1984; Breiman, 2001; Strobl et al.,



**FIGURE 5 |** Conditional permutation variable importance for the random forest with all background predictors for children's accuracy on two gender-assignment tasks. Predictors to the right of the 0.00 mark are significant.

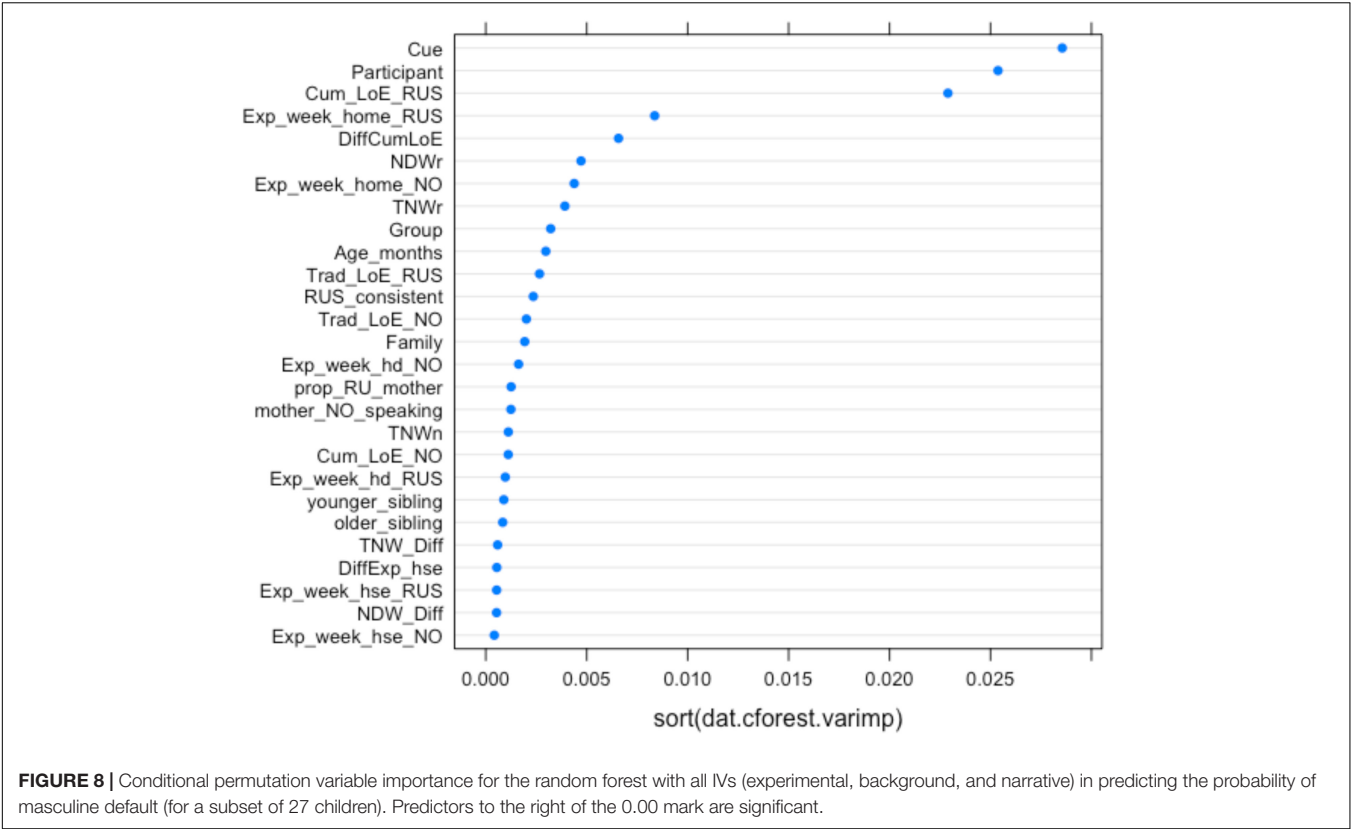


**FIGURE 6 |** Conditional permutation variable importance for the random forest with all background predictors for the probability of masculine default on two gender-assignment tasks. Predictors to the right of the 0.00 mark are significant.



**FIGURE 7 |** Conditional permutation variable importance for the random forest with all IVs (experimental, background and narrative) predicting HSs accuracy (for a subset of 27 children). Predictors to the right of the 0.00 mark are significant.





**TABLE 3 |** Background variables.

	All bilinguals	RR bilinguals	NR bilinguals	Russian L1
Number (N boys)	54 (27)	26 (15)	28 (13)	107 (67)
Age	4;0 – 10;2 (6;9) SD = 1.7	4;0 – 10;2 (6;9) SD = 1.6	4;3 – 9;9 (6;9) SD = 1.7	3;0 – 7;0 (5;2) SD = 1.1
Group (daycare/school)	22/32	10/16	12/16	107/0
CLoE R (in years)	0.66 – 7.39 (3.48) SD = 1.75	1.62 – 7.39 (4.68) SD = 1.45	0.66 – 5.78 (2.39) SD = 1.19	
CLoE N (in years)	0.33 – 7.03 (2.92) SD = 1.61	0.33 – 3.18 (1.79) SD = 0.67	1.49 – 7.03 (3.94) SD = 1.52	
Diff CLoE (N-R)	–5.54 – 6.25 (–0.56) SD = 2.89	–5.54 – 1.31 (–2.89) SD = 1.57	–3.43 – 6.25 (1.56) SD = 2.07	
Current exposure home/school/extra R (in %)	14 – 67 (42) SD = 16	3 – 67 (54) SD = 10	14 – 57 (30) SD = 12	
Current exposure home/school/extra N	33 – 86 (58) SD = 16	33 – 70 (46) SD = 10	43 – 86 (69) SD = 12	
Diff Current Exposure home/school/extra (N-R)	–34 – 72 (16) SD = 32	–34 – 40 (–8) SD = 19	–14 – 72 (38) SD = 24	
Input consistency R (Y/N)	28/26	19/7	10/18	
Older sibling (Y/N)	28/26	14/12	12/16	
Younger sibling (Y/N)	18/36	8/18	10/18	

2009; Tagliamonte and Baayen, 2012). Random forests analysis is a non-parametric non-linear statistical method which makes it possible to analyze complex interactions between a large number of variables (Baayen, 2008). A random forest is a so-called “ensemble of classification or regression trees (CARTs), where each tree in the ensemble is built according to the principle of recursive partitioning, where the feature space is recursively split into regions containing observations with similar response values” (Strobl et al., 2009, p. 324). The advantages of this method include its applicability to data that are not normally distributed,

as well as the fact that it allows for an automatic assessment of the relative importance of various variables in predicting the distribution of the data (cf., Tagliamonte and Baayen, 2012; Baayen et al., 2013).

However, as noted by Strobl et al. (2009), there are certain pitfalls connected to the fact that random forests were not developed in a stringent statistical framework, which might lead to potential confusion in the interpretation of main effects and interactions. To avoid these potential pitfalls, we decided to additionally run a standard mixed effects logistic regression analysis. We report the results of the models in turn and discuss the outcome of the analysis in the second part of the section.

## I. Random Forests

We fit two random forests models<sup>4</sup> (Hothorn et al., 2006; Strobl et al., 2008) to estimate the effect of 20 background variables (see above) on the children's accuracy with respect to gender assignment (Model 1), and on the probability of making masculine default errors (Model 2). Note that models of this type do not differentiate between fixed and random effects; thus, we also included the variable Participant to estimate the variance attributed to individual differences.

**Figures 5, 6** depict the relative importance of the predictors, using conditional permutation-based variable importance (see Strobl et al., 2008). The variables presented in **Figures 5, 6** appear in accordance with their relative importance as predictors of the children's accuracy (**Figure 5**) and probability of using masculine in non-masculine conditions (**Figure 6**). As the graphs show, the Participant is the most important predictor. This is not surprising, given that significant variability tied to the effect of individual participants is typical of psycholinguistic research in general (see e.g., Baayen, 2008; Tagliamonte and Baayen, 2012). The next most important predictor is CLoE to Russian, which is considerably more important than all other background variables. Significant predictivity is also detectable for Exposure to Russian and Norwegian at home per week (at present), Consistency of Russian input, Proportion of Russian with mother, Traditional length of exposure to Russian, Group, Presence of an older sibling, followed by the remaining variables. Note that Family type is generally ranked low in the hierarchy of predictors, suggesting that although the effect of Family type is significant, other variables have a much larger predictive power. In the next section, we present the analysis couched within the generalized linear mixed model approach.

## II. Generalized Linear Mixed Models

Recall that the reasons for including mixed effects logistic regressions were the following: (1) to assess the significance of the variables using a stringent statistical framework; (2) to assess whether the correlation between the variables is positive or negative; (3) to check for collinearity of fixed effects, and (4) to include random effects of Items and Participants (note that in the random forests approach, random and fixed effects are

not distinguished). In the logistic regressions, we included the ten most important variables from the random forests analysis and used them as predictors (apart from Participant, which was included as a random effect).

**Model 1:** Accuracy as predicted by the child's background. The following variables correlated significantly with the children's accuracy:

- (i) CLoE to Russian (positive correlation  $p < 0.001$ : children with a higher cumulative exposure index were more likely to assign gender accurately);
- (ii) Consistency of input in Russian (positive correlation  $p < 0.01$ : children who received consistent input in Russian were more accurate)
- (iii) The presence of an older sibling (negative correlation  $p < 0.001$ : children who had an older sibling were overall less accurate than those who did not).
- (iv) Group (positive correlation  $p < 0.05$ : schoolchildren were more accurate than daycare children).

**Model 2:** Defaulting to masculine agreement (in non-masculine conditions) as predicted by the child's background. The probability of using masculine as the default correlated significantly with the following predictors:

- (i) CLoE to Russian (negative correlation  $p < 0.001$ : children with a higher cumulative exposure index were less likely to default to masculine);
- (ii) Consistency of input in Russian (negative correlation  $p < 0.05$ : children who received consistent input in Russian were less likely to default to masculine);
- (iii) The presence of an older sibling (positive correlation  $p < 0.001$ : children who had an older sibling were more likely to make masculine default errors).

## Narrative Proficiency Measures

We collected child narratives in Norwegian and Russian with a subset of 27 out of the 54 participants. This group of 27 children is representative of the whole set of bilingual participants both in terms of family type (14 of the children were from NR homes and 13 from RR homes) and in terms of age (age range 4;0–9;6, mean age 6;8), which is comparable with the distribution in the whole sample.

Based on the children's narratives we calculated the following four variables: Total number of words in the Russian narrative (TNW), Total number of words in the Norwegian narrative (TNWn), Number of different words in the Russian narrative (NDW), and Number of different words in the Norwegian narrative (NDWn). In the analysis, we also included two relative variables: the difference between NDW in Norwegian and Russian, and the difference between the TNW in Norwegian and Russian (see **Table 4** for the descriptive statistics of the narrative variables). As evident from **Table 4**, the relative variables are mostly positive, suggesting that the majority of the participants used more words overall, as well as more different lexical words, in the Norwegian narratives than in the Russian narratives (only two children had slightly higher NDW and TNW scores in Russian than in Norwegian).

<sup>4</sup>The function *cforest* of the R package *party* was used for the analysis of variable importance (Breiman et al., 2006; cf. Strobl et al., 2008).

**TABLE 4 |** Narrative variables.

	All bilinguals	RR bilinguals	NR bilinguals
NDW R	4 – 53 (26) SD = 10	21 – 53 (31) SD = 8.9	4 – 37 (21) SD = 8.8
NDW N	27 – 74 (42) SD = 10	30 – 63(43) SD = 10	27 – 74 (41) SD = 11
Diff NDW (N-R)	–12 – 41 (16) SD = 11	–12 – 32 (12) SD = 12	4 – 41 (20) SD = 8
TNW R	8 – 88 (42) SD = 17	32 – 88 (51) SD = 16	8 – 54 (34) SD = 13
TNW N	47 – 180 (90) SD = 29	54 – 126 (89) SD = 23	47 – 180 (90) SD = 34
Diff TNW (N-R)	–1 – 137 (48) SD = 27	–1 – 73 (38) SD = 23	32 – 137 (56) SD = 28

Based on these measures we conducted a combined analysis of the data from the 27 bilingual children which included all variables: experimental variable (Condition), individual-level variable Participant, and all background and narrative variables.

**Figures 7, 8** show the relative importance of the predictors in explaining accuracy (**Figure 7**) and probability of defaulting to masculine (**Figure 8**).

As **Figures 7, 8** show, the four most important predictors are Participant, Condition (Cue), CLoE to Russian, and NDW in the Russian narrative (our lexical diversity measure in the heritage language). Other background and narrative predictors are ranked below these variables.

To test the effects of the predictors in a more stringent statistical approach, we fit a set of generalized linear mixed effects logistic regression models to predict accuracy (Models 1a and 2a) and probability of making masculine default errors (Models 1b and 2b) with a combined set of experimental, background, and narrative predictors. Models 1a and 1b included the following main effects: an experimental variable (Condition) and three BiLEC variables (CLoE to Russian, presence of an older sibling, and weekly exposure to Norwegian, i.e., the variables which were ranked highest in predicting masculine default errors in the random forests analysis). Models 2a and 2b additionally included two narrative variables (NDW in the Russian and Norwegian narratives). We excluded other narrative variables to avoid collinearity and achieve model convergence. Participants and Items were taken as random effects in both models. The aim of this analysis was to establish whether the inclusion of the narrative variables would significantly improve model fit, or whether the narrative variables can be safely disregarded.

The two main predictors: CLoE to Russian and NDW in the Russian narrative both correlated positively with the children's accuracy ( $p < 0.05$  and  $p < 0.01$ , respectively) and negatively with the probability of masculine default errors ( $p < 0.05$  for both correlations). NDW in the Norwegian narrative was not a significant predictor.

A likelihood ratio test (ANOVA) of Models 1a and 2a as well as Models 1b and 2b showed that the bigger models (which additionally included predictors from the narrative

tasks) should be preferred, despite the higher number of predictors involved. Models 2a and 2b were significantly better than their Model 1 counterparts ( $p < 0.001$  in both cases).

To sum up, converging results from parametric and non-parametric statistical modeling show that the number of different words (NDW) used in the Russian narrative task and the CLoE to Russian (CumLoE) are both reliable predictors of children's grammatical development in their heritage language, illustrated by the acquisition of gender assignment patterns in this study. Furthermore, these predictors do not overlap, but complement each other, and the inclusion of the narrative measure significantly improves model fit and explains more variance in the data.

Finally, we also tested whether the inclusion of the dominance variable (differences in the cumulative amount of exposure between Norwegian and Russian) that was ranked high in the random forests analysis could significantly improve the model fit. We fit two additional generalized linear mixed effects regression models, Models 3a and 3b, which included the same predictors as Models 2a and 2b, but additionally a variable reflecting Difference in CLoE between Norwegian and Russian. A likelihood ratio test (ANOVA) of Models 2a and 3a and Models 2b and 3b showed that the bigger models (which additionally include the dominance predictor) do not survive the comparison. The difference between the models was not significant, and the criteria for model selection (Bayesian information criterion and Akaike information criterion) were smaller for Models 2a and 2b, which is preferable. This means that adding the relative dominance variable to the model does not improve model fit and does not explain any additional part of the variance that is not covered by the absolute exposure and proficiency variables, such as CLoE and lexical diversity in the heritage language.

## DISCUSSION

### Gender Knowledge and Cue Sensitivity

With regard to gender knowledge and cue sensitivity of bilinguals we asked the following two main questions:

- (1) Do heritage speakers of Russian (HSs) differ significantly from monolinguals and in what conditions (on the real and nonce word tasks)?
- (2) Are there any differences between nouns with transparent and nouns with non-transparent morphophonological gender cues, for both monolinguals and HSs (on the real and nonce word tasks)?

In the real word task, no differences between groups were found in the masculine conditions, with RR, NR and monolingual children all performing at ceiling. However, with masculine being considered the default gender, target-like performance in the masculine condition does not allow us to disambiguate between actual internalized knowledge of the grammatical gender of a given item and a defaulting strategy. In the feminine and neuter conditions, the statistical analysis revealed that the NR children were significantly less target-like than monolingual

controls, while no statistical difference was found between the monolinguals and the RR children. Overuse of masculine in non-masculine conditions was found in both bilingual groups, but it was most prevalent in the data of the NR bilinguals. In the case of vowel-final items (feminine transparent, neuter transparent, neuter opaque), these results are different from those of Polinsky (2008), who finds that vowel-final items are overgeneralized to the feminine and consonant-final items to the masculine in adult heritage speakers. In other words, Polinsky (2008) finds evidence for a restructured (and simplified) grammatical gender system, where the binary opposition in grammatical gender (masculine vs. feminine) is marked by means of a binary phonological contrast (consonants vs. vowels).

Notably, our participants and the participants in Polinsky's study differ in at least two important respects: age and majority language. The participants in Polinsky's study are adults who were born and raised in Russian immigrant families in the United States. Our participants were children born and raised in Russian-Russian and Norwegian-Russian families in Norway. Given significant age differences between the groups, we cannot rule out the possibility that the two distinct patterns (defaulting to masculine across the board vs. restructuring of the 3-gender system into a simplified masculine vs. feminine system) might reflect two different stages of heritage language development (cf. Polinsky, 2016). For instance, it may be the case that with more exposure to Russian, feminine and neuter genders will be acquired by some of our participants and a three-gender system will be developed. However, it may also be the case that some of the children will never acquire the target grammatical gender distinctions, and hence this can be considered a case of incomplete acquisition (in fact, we believe that for two of our participants who are already above the age of 8 but have produced exclusively masculine agreement in both gender tasks and narratives, this might be the case). The participants in Polinsky (2008) study may on the other hand already show signs of attrition – it is possible that they had had a three-gender system at some point of their grammatical development (e.g., at the pre-school age), but later developed a simplified two-gender system due to attrition and lack of contact with Russian.

Cross-linguistic influence from the majority language may also be a factor contributing to the observed difference between the patterns reported in Polinsky (2008) and in our study. Masculine is without doubt the morphological default in Norwegian, with feminine being the most vulnerable gender. In many dialects, feminine is disappearing and is being replaced by masculine (Lødrup, 2011; Rodina and Westergaard, 2015; Busterud et al., forthcoming). It is conceivable that the role of masculine in heritage Russian is strengthened under the influence of Norwegian, a language with a strong masculine default. Potential supporting evidence for this idea comes from a study concerning gender acquisition in Russian by Russian-German bilingual children (Dieser, 2009). This study found that Russian-German children, especially those with a small CLoE to Russian and a low amount of different words in Russian narratives, tended to default to feminine, which has been argued to be the default gender in child German (Kupisch et al., 2018).

Interestingly, we do find that RR children overgeneralize neuters to the feminine as well as to masculine (and the trend persists with nonce items as well). This might be taken as evidence that RR children are more likely to assign nouns to gender categories based on (some generalizations over) morphophonological cues, while NR children simply rely on the default. In other words, the RR children (as well as the monolinguals) seem to be more sensitive to the fact that there is a phonological similarity between final-unstressed neuters and final-unstressed feminines, and their mis-assignment of the former stems from this knowledge. In contrast, NR children prevalently overgeneralized final-unstressed neuters to the masculine, which suggests that they are oblivious to this similarity.

Finally, the transparency of gender cues only played a significant role in the feminine conditions, with all groups of participants showing higher accuracy with transparent feminine items. It is likely that the distinction between transparent and opaque masculines may be masked by the masculine default strategy. The difference between transparent and opaque neuters did not surface in the case of bilinguals, because neuters were generally a challenge for them due to their low input frequency.

Turning now to the nonce word task, it was found that both RR and NR bilinguals were significantly less target-appropriate than monolingual controls on the feminine transparent and neuter transparent nouns. The difference in accuracy between the two bilingual groups reached statistical significance in the feminine transparent condition, although the NR bilinguals gave more non-target appropriate responses than the RR bilinguals in the neuter transparent condition as well. The RR bilinguals patterned with monolinguals in that they preferred feminine in the ambiguous FN and FM conditions, while the NR bilinguals used significantly more masculine agreement in both conditions. Overall, the results show that purely cue-based gender assignment is more challenging for the bilinguals, while the differences between the bilingual groups indicate that the amount of exposure plays a role. At the same time, it needs to be stressed that all groups of participants showed sensitivity to phonological gender cues – albeit to different degrees. Neuter responses were given exclusively in neuter conditions, and the proportion of feminine responses was significantly higher in those conditions where it is target-appropriate.

Our research questions 3 and 4 addressed the role of lexical learning. Specifically, we asked whether bilinguals rely on lexically stored gender features and whether this behavior may be more pronounced in HSs than in monolinguals. Our results show that accuracy across the three transparent conditions on the real word task was significantly higher than on the nonce word task for all three groups of participants, with the interaction of group and task being not significant. This might be taken as evidence that lexical learning of the gender category of familiar nouns in addition to cue-based assignment is an important strategy in grammatical gender acquisition for both bilinguals and monolinguals. Additionally, the difference between real and nonce items may be attributed to the difference



in the cognitive load required by each task. It is reasonable to assume that when the grammatical gender of a noun is acquired (regardless of whether it was deduced from noun-internal cues or distributional information) it is stored in the lexical entry and retrieved as needed rather than computed online each time the lexical item is invoked (Caramazza et al., 1988). On-the-go gender assignment of a novel noun, on the other hand, presupposes online computation, which is arguably a more cognitively demanding process than retrieval and therefore more error-prone.

## The Role of Exposure, Proficiency, and Dominance

Research questions 5–8 investigated to what extent a composite measure of proficiency and amount of exposure influenced heritage speakers' performance on gender assignment tasks. Specifically, we asked which of the background variables were the most reliable predictors of performance on the gender assignment tasks, whether lexical measures of bilinguals' performance in the narrative task would correlate significantly with their performance on the gender assignment tasks, and whether the use of a composite measure of background data and narrative proficiency would have an advantage in predicting the performance of HSSs on the gender assignment task (as compared to using only background or only proficiency measures).

Question 5 asked which of the 20 background variables were the most reliable predictors of bilinguals' performance on the gender assignment tasks. The results showed that CLoE to Russian was by far the most reliable and predictive variable that accounted for the largest portion of the variance in the data. A number of other variables, such as Consistency of input in Russian, Traditional length of exposure to Russian, Proportion of Russian with the Russian-speaking parent (mother), as well as Presence/absence of an older sibling were also high in the hierarchy of the most important predictors (see **Figures 5, 6**). Statistical significance of CLoE to Russian (positive correlation with accuracy, negative correlation with the probability of defaulting to M) as well as a number of other background variables (see above) was confirmed through subsequent generalized linear mixed effects regression analysis. To sum up, converging results of parametric and non-parametric statistical models indicate that CLoE to Russian is the most robust and reliable background variable that can be taken as a proxy of the children's amount of exposure to Russian. In other words, our results show that CLoE to Russian is a better predictor of heritage speakers' level of acquisition of grammatical gender in Russian than other background variables.

Questions 6 and 7 asked whether the lexical measures of proficiency obtained from the narrative samples in both languages correlated with HSS' performance on the gender assignment tasks. Our analysis included four variables: Total number of words (TNW) in the Russian narrative, TNW in the Norwegian narrative, Number of different words (NDW) in the Russian narrative, and NDW in the Norwegian narrative. NDW in the Russian narrative was ranked highest of all

narrative variables in the random forest analysis. Subsequent generalized mixed logistic regression analysis confirmed that out of all narrative variables, only NDW in the Russian narrative correlated significantly with accuracy (positive correlation) and with the probability of masculine default errors (negative correlation).

Multiple recent studies have suggested that combining various language proficiency measures (production, comprehension, repetition, etc.) with background measures quantifying language exposure and use would be fruitful in modeling heritage speakers' grammatical abilities (Montrul, 2015, 2016, chapter 6; Polinsky, 2018: chapter 3). We included two types of variables into our analysis and showed that background measures and narrative proficiency measures are both significant predictors of children's performance on gender assignment tasks. Furthermore, we showed that combining background measures with narrative proficiency measures improved the predictive power of the statistical model. This indicates that narrative proficiency measures in the heritage language have an independent value as predictors of HSSs' acquisition of grammatical gender, in addition to language exposure variables.

Our last question addressed the independent effect of language dominance – in addition to absolute background and narrative proficiency measures – on the HSSs' performance with respect to gender assignment. We used three variables to quantify dominance: (i) the difference in the CLoE between the majority and the minority languages, (ii) the difference in current exposure to the two languages, and (iii) the difference in the relative scores on the narrative tasks in the two languages. The only variable that was ranked relatively high on the variable importance hierarchy in the random forests analysis – although still below CLoE to Russian – was the difference in the cumulative exposure between the two languages. However, the results of a model comparison showed that the inclusion of this variable in addition to CLoE to Russian did not improve model fit. It is thus likely that the importance of this variable might be an artifact of collinearity between this variable and CLoE to Russian: the more Russian input the children accumulate, the smaller the difference between cumulative exposure to Norwegian and to Russian. None of the dominance variables turned out to have an independent value for our analysis, since all the variation they accounted for could be captured by variables that measure absolute cumulative exposure and proficiency in the heritage language.

## CONCLUSION

In this paper, we have shown that a combination of background variables and proficiency measures predicts heritage speakers' performance on grammatical gender tasks in Russian better than background measures or narrative proficiency measures taken in isolation. We carried out two production experiments investigating gender assignment to real as well as nonce words in Russian, including all three genders and transparent as well as opaque cues. Participants were 54 Norwegian-Russian bilingual children living in Norway (age range 4;0–10;2) and 107 monolingual controls. Background information was collected

through the *Bilingual Language Exposure Calculator* (Unsworth, 2013), and the proficiency measure was based on the MAIN semi-spontaneous narratives (Gagarina et al., 2012). As many as 20 background variables and six proficiency measures were included in the statistical analysis of the participants' performance on the gender tasks. We also included three dominance variables (the difference in CLoE between the majority and the heritage language, the difference in current exposure to the two languages, and the difference in scores on Russian and Norwegian narratives). The best predictors turned out to be a combination of three background variables (CLoE to Russian, consistency of input, and the presence of older siblings) and one proficiency measure, lexical diversity as defined by the number of different words in the Russian narrative. Interestingly, our statistical analysis showed that the dominance variables are not robust predictors for the bilingual children's performance on gender assignment. We argue that these results support Montrul's (2015) distinction between dominance and proficiency: Language dominance vs. non-dominance is a relative concept and may reflect considerable variation with respect to proficiency in the heritage language.

## ETHICS STATEMENT

The project was registered and approved by the Norwegian Social Science Data Service (NSD, <http://www.nsd.uib.no>). Data collection was conducted in accordance with NSD's ethical principles. Written informed consent was obtained from parents of all the participants prior to testing.

## REFERENCES

- Au, T. K., and Romo, L. F. (1997). "Does childhood language experience help adult learners?" in *The Cognitive Processing of Chinese and Related Asian Languages*, ed. H.-C. Chen (Hong Kong: Chinese University Press), 417–441.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511801686
- Baayen, R. H., Janda, L. A., Nessel, T., Endresen, A., and Makarova, A. (2013). Making choices in Slavic: pros and cons of statistical methods for rival forms. *Russian Linguist.* 37, 253–291. doi: 10.1007/s11185-013-9118-6
- Bedore, L. M., Peña, E., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., et al. (2012). The measure matters: language dominance profiles across measures in Spanish–English bilingual children. *Biling. (Camb. Engl.)* 15, 616–629. doi: 10.1017/S1366728912000090
- Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *Int. J. Biling.* 14, 422–446.
- Bohman, T. M., Bedore, L. M., Peña, E. D., Méndez-Pérez, A., and Gillam, R. B. (2010). What you hear and what you say: language performance in Spanish–English bilinguals. *Int. J. Biling. Educ. Biling.* 13, 325–344. doi: 10.1080/13670050903342019
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Breiman, L., Cutler, A., Liaw, A., and Wiener, M. (2006). Breiman and Cutler's Random Forests for Classification and Regression. R Package Version 4.5–16.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. New York, NY: Chapman and Hall.

## AUTHOR CONTRIBUTIONS

All co-authors are responsible for the conception of the work and experimental design. NM carried out the collection of control data with monolingual Russian children. NM, YR, and OU carried out the collection of data with bilingual participants. MW collected Norwegian narrative data with bilingual participants. NM is responsible for data analysis and interpretation of the results. All authors share responsibility for drafting of the work and final approval of the version to be published.

## FUNDING

This research was supported by a grant from the Research Council of Norway for the project MiMS (Micro-variation in Multilingual Acquisition and Attrition Situations), MW, project number 250857. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

## ACKNOWLEDGMENTS

We thank the participating families and schools for their cooperation, Isabel Nadine Jensen for data collection and transcription of the Norwegian narratives, and Anna Afanasyeva for the transcription of the Russian narratives.

- Busterud, G., Lohndal, T., Rodina, Y., and Westergaard, M. (forthcoming). The loss of feminine gender in Norwegian: a dialect comparison. *J. Comp. Germanic Linguistics*.
- Caramazza, A., Laundanna, A., and Romani, C. (1988). Lexical access and inflectional morphology. *Cognition* 28, 297–332. doi: 10.1016/0010-0277(88)90017-0
- Ceitlin, S. N. (2005). "Kategorija roda v detskoj rechi [The category of gender in child language]," in *Problemy funkcionalnoj grammatiki [Issues in Functional Grammar]*, eds A. V. Bondarenko, and S. A. Shubik (St Petersburg: Rossijskaja Akademija Nauk), 346–375.
- Ceitlin, S. N. (2009). *Očerki po slovoobrazovaniju i formoobrazovaniju v detskoj rechi. [On Inflection and Derivation in Child Language]*. Moscow: Znak.
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguist. Approaches Biling.* 1, 318–342. doi: 10.1075/lab.1.3.05cho
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166119
- Corbett, G. (2007). "Gender and noun classes," in *Language Typology and Syntactic Description*, ed. T. Shopen (Cambridge: Cambridge University Press), 241–279.
- De Cat, C., and Serratrice, L. (2017). Predicting language proficiency in bilingual children. *PsyArXiv [Preprint]*. doi: 10.31234/osf.io/btjy4
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Appl. Psycholinguist.* 28, 411–424. doi: 10.1017/S0142716407070221
- Dieser, E. (2009). *Genuserwerb im Russischen und Deutschen. Korpusgestützte Studie zu ein- und zweisprachigen Kindern und Erwachsenen*. München: Sagner. doi: 10.3726/b12030

- Flhannchadha, S. N., and Hickey, T. M. (2017). Acquiring an opaque gender system in Irish, an endangered indigenous language. *First Lang.* 37, 475–499. doi: 10.1177/0142723717702942
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., et al. (2012). *Multilingual Assessment Instrument for Narratives (MAIN)*. ZAS Papers in Linguistics. Berlin: ZAS.
- Gathercole, V. C. M., and Thomas, E. M. (2005). "Minority language survival: input factors influencing the acquisition of Welsh," in *Proceedings of the 4th International Symposium on Bilingualism*, eds J. Cohen, K. McAlister, K. Rolstad, and J. MacSwan (Somerville, MA: Cascadia Press), 852–874.
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingualism* 12, 213–237. doi: 10.1017/S1366728909004015
- Gutiérrez-Clellen, V. F., and Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Appl. Psycholinguist.* 24, 267–288. doi: 10.1017/S0142716403000158
- Gvozdev, A. N. (1961). *Formirovanie u rebenka grammatičeskogo stroja russkogo jazyka*. [Language Development of a Russian Child]. Moscow: APN RSFSR.
- Hall, K. C., Blake, A., Fry, M., Mackie, S., and McAuliffe, M. (2016). *Phonological CorpusTools, Version 1.2*. [Computer Program].
- Horst, J. S., and Hout, M. C. (2016). The novel object and unusual name (noun) database: a collection of novel images for use in experimental research. *Behav. Res. Methods* 48, 1393–1409. doi: 10.3758/s13428-015-0647-3
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics* 7, 355–373. doi: 10.1093/biostatistics/kxj011
- Iosad, P. (2012). Vowel reduction in Russian: no phonetics in phonology. *J. Linguist.* 48, 521–571. doi: 10.1017/S0022226712000102
- Kupisch, T., Geiß, M., and Natalia, M. (2018). *Are Bilinguals Sensitive to Morpho-Phonological Gender cues in German? Poster Presented at Theoretical and Experimental Approaches to Gender*. Available at: <https://sites.google.com/site/thegen2018/talks-and-program>
- Lenth, R. V. (2016). Least-squares means: the r package lsmeans. *J. Stat. Softw.* 69, 1–33. doi: 10.18637/jss.v069.i01
- Lødrup, H. (2011). Hvor mange genus er det i Oslo-dialekten? [how many genders are there in the Oslo dialect?]. *Maal Og Minne* 2, 120–136.
- Montrul, S. A. (2015). "Dominance and proficiency in early and late bilingualism," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35.
- Montrul, S. A. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139030502
- 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., Nicoladis, E., and Crago, M. (2007). "French-English bilingual children's acquisition of the past tense," in *Proceedings of the 31st Annual Boston University Conference on Language Development*, eds H. Caunt-Nulton, S. Kulatilake, and I.-H. Woo (Somerville, MA: Cascadia Press), 497–507.
- Place, S., and Hoff, E. (2011). Properties of dual language exposure that influence two-year-olds' bilingual proficiency. *Child Dev.* 82, 1834–1849. doi: 10.1111/j.1467-8624.2011.01660.x
- Polinsky, M. (2008). Gender under incomplete acquisition: heritage speakers' knowledge of noun categorization. *Herit. Lang. J.* 6, 40–71.
- Polinsky, M. (2016). Bilingual children and adult heritage speakers: the range of comparison. *Int. J. Biling.* 22, 547–563. doi: 10.1177/1367006916656048
- Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press. doi: 10.1017/9781107252349
- Popova, M. I. (1973). "Grammatical elements of language in the speech of pre-preschool children," in *Studies of Child Language Development*, eds C. A. Ferguson and D. I. Slobin (New York, NY: Holt, Rinehart and Winston), 269–280.
- Rodina, Y. (2017). Narrative abilities of preschool bilingual Norwegian-Russian children. *Int. J. Biling.* 21, 617–635. doi: 10.1177/1367006916643528
- Rodina, Y., and Westergaard, M. (2013). "Two gender systems in one mind: the acquisition of grammatical gender in Norwegian-Russian bilinguals," in *Multilingualism and Language Diversity in Urban Areas: Acquisition, Identities, Space, Education*, eds P. Siemund, I. Gogolin, M. E. Schulz, and J. Davydova (Amsterdam: John Benjamins), 95–126.
- Rodina, Y., and Westergaard, M. (2015). Grammatical gender in Norwegian: language acquisition and language change. *J. Germanic Linguist.* 27, 145–187. doi: 10.3389/fpsyg.2016.00344
- Rodina, Y., and Westergaard, M. (2017). Grammatical gender in bilingual Norwegian-Russian Acquisition: the role of input and transparency. *Bilingualism* 20, 197–214. doi: 10.1017/S1366728915000668
- Schmeißer, A., Hager, M., Gil, L., Jansen, V., Geveler, J., Eichler, N., et al. (2015). "Related but different: the two concepts of language dominance and language proficiency," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 36–65.
- Schwartz, M., Minkov, M., Dieser, E., Protassova, E., Moin, V., and Polinsky, M. (2015). Acquisition of Russian gender agreement by monolingual and bilingual children. *Int. J. Biling.* 19, 726–752. doi: 10.1177/1367006914544989
- Sharoff, S. (2002). "Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics," in *Proceedings of the Language Resources and Evaluation Conference (LREC02)*, Las Palmas.
- Sorace, A. (2005). "Selective optionality in language development," in *Syntax and Variation: Reconciling the Biological and the Social*, eds L. Cornips and K. Corrigan (Amsterdam: John Benjamins), 55–80. doi: 10.1075/cilt.265.04sor
- Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics* 9:307. doi: 10.1186/1471-2105-9-307
- Strobl, C., Tutz, G., and Malley, J. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323–348. doi: 10.1037/a0016973
- Tagliamonte, S. A., and Baayen, R. H. (2012). Models, forests and trees of York English: was/were variation as a case study for statistical practice. *Lang. Var. Change* 24, 135–178. doi: 10.1017/S0954394512000129
- Treffers-Daller, J., and Korybski, T. (2015). "Using lexical diversity measures to operationalise language dominance in bilinguals," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 106–123.
- 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. (2015). "Amount of exposure as a proxy for dominance in bilingual language acquisition," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173.
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., and Tsimpli, I. (2014). On 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
- Zakharova, A. V. (1973). "Acquisition of forms of grammatical case by preschool children," in *Studies of Child Language Development*, eds C. A. Ferguson and D. I. Slobin (New York, NY: Holt, Rinehart and Winston), 281–284.

**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 Mitrofanova, Rodina, Urek and Westergaard. 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.



# Case Marking in Hindi as the Weaker Language

Silvina Montrul<sup>1\*</sup>, Archana Bhatia<sup>2</sup>, Rakesh Bhatt<sup>1</sup> and Vandana Puri<sup>1</sup>

<sup>1</sup> Department of Spanish and Portuguese/Department of Linguistics, University of Illinois at Urbana-Champaign, Urbana, IL, United States, <sup>2</sup> Florida Institute for Human and Machine Cognition, Pensacola, FL, United States

Does language dominance modulate knowledge of case marking in Hindi-speaking bilinguals? Hindi is a split ergative language with a rich morphological case system. Subjects of transitive perfective predicates are marked with ergative case (-ne). Human specific direct objects, indirect objects, and dative subjects are marked with the particle -ko. We compared knowledge of case marking in Hindi–English bilinguals with different dominance patterns: 23 balanced bilinguals and two groups of bilinguals with Hindi as their weaker language: 24 L2 learners of Hindi with age of acquisition (AoA) of Hindi in adulthood and 26 Hindi heritage speakers with AoA of Hindi since birth in oral production and acceptability judgments. The balanced bilinguals outperformed the English-dominant bilinguals; the L2 learners and the heritage speakers, who showed similar lower command of the Hindi case marking system, with the exception of -ko marking as a function of specificity with direct objects. We consider how dominant language transfer, AoA of Hindi, and input factors may explain the acquisition and knowledge of morphology in Hindi as the weaker language.

**Keywords:** Hindi, dominance, heritage speakers, second language, case, ergativity, differential object marking

## OPEN ACCESS

### Edited by:

Esther Rinke,  
Goethe-Universität Frankfurt am Main,  
Germany

### Reviewed by:

Niharika Singh,  
Centre of Behavioural and Cognitive  
Sciences, India  
Maria Pilar Larrañaga,  
Universität Hamburg, Germany

### \*Correspondence:

Silvina Montrul  
montrul@illinois.edu

### Specialty section:

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

**Received:** 13 March 2018

**Accepted:** 15 February 2019

**Published:** 19 March 2019

### Citation:

Montrul S, Bhatia A, Bhatt R and  
Puri V (2019) Case Marking in Hindi  
as the Weaker Language.  
Front. Psychol. 10:461.  
doi: 10.3389/fpsyg.2019.00461

## INTRODUCTION

Bilinguals know two or more languages and may use them to different degrees. Although the term *bilingual* continues to conjure stable and equally highly proficient linguistic knowledge and use of two languages, the reality is that most bilinguals have unequal command of the two languages overall, by language skills, and in specific linguistic domains. Dominance is the relative weight and relationship of the two languages of a bilingual in terms of language use and degree of proficiency (Silva-Corvalán and Treffers-Dallers, 2016), with the two languages having relatively similar strength, or one being stronger/weaker than the other. Factors that contribute to language dominance may include age of acquisition (AoA) or age of bilingualism, estimations of language input, degree of language use, and proficiency in each language (Montrul, 2016a). Bilingual balance or imbalance may be a reflection of the Complementarity Principle (Grosjean, 2008): the idea that bilinguals use their languages in different situations and for different purposes along the lifespan.

Does language dominance modulate knowledge of specific structural properties of a language in bilinguals? How does AoA and context of learning affect the acquisition of a weaker language? Does acquisition of a language very early in a naturalistic setting always have a long-term advantage? We answer these questions by looking at the linguistic situation depicted in **Figure 1**: we compare the linguistic abilities of Hindi–English bilinguals with different patterns of dominance (balanced vs. unbalanced bilinguals), and within the unbalanced groups we include bilinguals who share the same dominance pattern—English is the dominant language and Hindi is the weaker



language—but differ in their AoA of their weaker language and in the context of learning: heritage speakers of Hindi and second language (L2) learners of Hindi in the United States.

For L2 learners the weaker language is the L2 but for heritage speakers it is the L1 [or language A (LA) if acquired simultaneously with language B (LB)]. L2 learners are late bilinguals because the L2 is typically acquired around or after puberty and in an instructed setting. In the case of heritage speakers—early bilinguals—their L1 is a minority language and it may become weak over time due to shift in childhood, becoming secondary in several domains of use. The majority language ends up being the dominant language in early adulthood and, if the heritage language is learned simultaneously with the majority language from birth, the heritage language can lag in development during the school-age period (Polinsky, 2006; Rothman, 2009; Carreira and Kagan, 2011; Montrul and Ionin, 2012; Montrul, 2016b). Yet, the weaker language in heritage speakers is considered a native language (Montrul, 2013; Rothman and Treffers-Daller, 2014; Bayram et al., 2017).

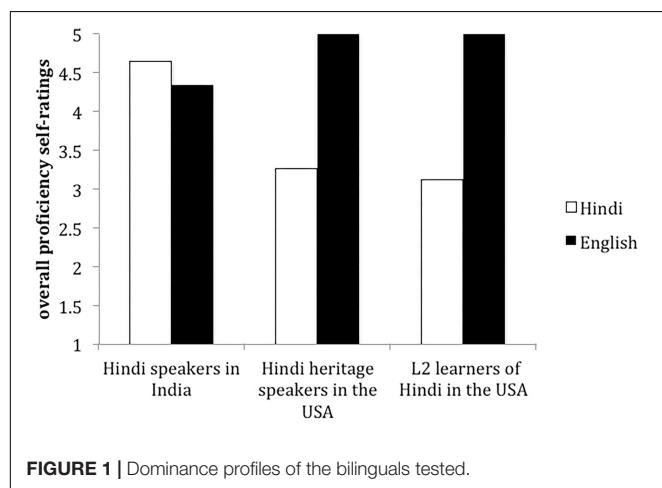
The heritage speakers and the L2 learners in our study are comparable in patterns of language dominance and in proficiency in the two languages, but differ in age and context of acquisition of the weaker language. We investigate whether these admittedly confounded variables (age and context acquisition of the weaker language) play a role in the morphological competence of bilinguals with similar dominance profiles. Given differences in timing, context and modality (auditory, written) of input experience, the question of whether and how heritage speakers and L2 learners differ in their linguistic knowledge continues to generate intense theoretical, empirical and practical interest (Santos and Flores, 2016; Perpiñán, 2017). Early acquisition and language experience gives heritage speakers a clear advantage compared to L2 learners when it comes to phonetics and phonology. Research has consistently shown that heritage speakers perform more native-like than L2 learners in phonological perception and production (Chang et al., 2011; Lukyanenko and Gor, 2011; Chang, 2016). However, similar advantages for heritage speakers over L2 learners in morphosyntax have been less consistent, and the

results in this area are more variable (Au et al., 2002; Foote, 2010). AoA is confounded with experience and context of learning in these two groups. Language experience, which includes amount and nature of input, is very relevant for the acquisition and mastery of morphology. Unsworth et al. (2014) attempted to disentangle the role of AoA from the role of input in Greek-English and Dutch-English bilingual children with different onsets of bilingualism and diverse language experiences and found a weak effect for AoA and a stronger effect of cumulative length of exposure with the acquisition of gender marking in nominals. Both L2 learners and heritage speakers are exposed to less input in the L2/heritage language than monolingually-raised children and balanced bilinguals who use the language more frequently and consistently. Yet heritage speakers may have more cumulative exposure to the weaker language than L2 learners because they were exposed to it earlier. If the acquisition of morphology is largely influenced by input factors, L2 learners and heritage speakers may show similar accuracy patterns on morphology. But if timing of input (i.e., AoA) determines the outcome of morphological acquisition instead, as it seems to impact phonology, heritage speakers may be more accurate with morphology than L2 learners. Because findings on morphology with respect to this question in adult heritage speakers and L2 learners have been inconclusive, more research on different languages and with different morphological patterns is warranted.

Our study contributes to this critical debate and is unique in examining morphology in Hindi, an understudied language. We focus on case morphology because case marking is very vulnerable to erosion in heritage languages in general (Benmamoun et al., 2013; Putnam and Sánchez, 2013; Kim et al., 2016), and is similarly difficult to acquire for L2 learners whose L1 does not mark case overtly (Papadopoulou et al., 2011; Baten and Verbeke, 2015; Baten et al., 2016). Hindi is a split-ergative language with a complex system of morphological case marking, and presents a challenging learning task for bilinguals whose dominant language has a nominative-accusative case pattern and does not mark morphological case overtly, like English. By probing into the syntactic and semantic distribution of the case particles *-ne* (ergative) and *-ko* (accusative, dative) in oral production and in an acceptability judgment task (AJT), we investigate how the complexity of morphological form-meaning mappings interacts with limited input factors and AoA in contributing to the acquisition of Hindi case morphology.

## Morphology in the Weaker Language

In both L2 acquisition and heritage language acquisition, morphology seems to be a bottleneck compared to other areas of the grammar (Slabakova, 2008; Montrul, 2018): it is difficult to acquire, easy to lose, and displays variability in L2 learners and heritage speakers. Following Distributed Morphology (Halle and Marantz, 1993), Lardiere (2009) explained how morphological knowledge may be acquired and computed during L2 learning. Language learners must assemble the lexicon of a language by associating lexical items (affixes, stems) with the specific formal features that the language selects from the inventory provided by Universal Grammar (Chomsky, 2001). Morphemes consist of



features and feature bundles that encode phonological, semantic (+ interpretable) and syntactic (– interpretable) information, and languages assemble and combine features such as [+/- wh], [+/- plural], [+/- definite] in lexical and functional categories in different ways. Learners must learn which morpholexical forms and their allomorphic variants express which specific syntactic and semantic features, as well as the contextual conditions under which such morphological forms are realized overtly or as zero marking. Our study considers three factors that contribute to how fast and accurately bilinguals whose weaker language is an L1 or an L2 correctly reassemble and reconfigure features in the target language: (1) dominant language transfer, (2) the nature of the morphology itself (i.e., complexity assembling semantic and syntactic features of morphemes, and the transparency of form-meaning mappings), and (3) AoA of the weaker language.

It is well-established that in L2 acquisition, the dominant L1 guides and constrains the acquisition of the L2 (Schwartz and Sprouse, 1996). Morphological variability often occurs if the learner has not acquired the relevant abstract features or their values, or a given feature may be part of their L1 grammars but the learner lacks the relevant knowledge of the conditions for expressing the given feature in the L2. L2 learners look for morpholexical equivalents of their dominant language (the L1) features in the L2, assuming initially that L2 values are the same as in their L1 (Montrul, 2001). Development proceeds when learners are able to determine how to assemble the lexical items of the L2, by reconfiguring the feature values in lexical items and functional categories from their L1 to those of the L2 in cases where these are different (Lardiere, 2009).

For example, a type of case marking, Differential Object Marking (DOM), is frequently omitted in L2 and heritage language acquisition (Montrul, 2011). In Spanish, animate and specific direct objects are obligatorily marked with the preposition “a” (*Juan vio a María* “Juan saw María”). In Turkish, specific direct objects are marked with the accusative affix *-(y)I* (*Ayşe adam-ı gördü* “Ayşe saw the man”), and we will see that Hindi is similar to Turkish. Let’s assume that Spanish DOM bundles two features in the lexical item “a” [+animate, +specific] whereas Turkish *-(y)I* bundles only one [+specific]. Turkish-speaking learners of Spanish (Montrul and Gürel, 2015) have been shown to be more successful than English-speaking learners of Spanish (Guijarro Fuentes, 2012) at acquiring the feature specifications of DOM in Spanish. This is because Turkish-speaking learners had to only add the new [+animate] feature of Spanish to their L2 Spanish representation, whereas the English-speaking learners had to build the representation for DOM in L2 Spanish anew, with its feature specification [+animate, +specific]. Like Spanish, Romanian also bundles the features [+animate, +specific] in the lexical item *pe* (the DOM marker in Romanian), and Romanian-speaking learners of Spanish have been shown to exhibit native-like acquisition of Spanish DOM (Montrul, in press). These studies provide clear evidence of L1 influence in the L2 acquisition of case morphology. As in L2 acquisition, transfer from the majority (stronger) language to the heritage (weaker) language is quite common in heritage speakers. The erosion and simplification of case found in Russian

(Polinsky, 2006), in Spanish (Montrul and Bowles, 2009) and in Korean heritage speakers (Kim et al., 2016) in the United States could partially be due to the fact that English does not mark morphological case.

In addition to dominant language influence or transfer, the syntactic and semantic composition of morphemes and feature assembly is another likely cause of difficulty in L2 learners and heritage speakers. Morphological complexity has been linked to the number of elements making up individual morphemes and morphological systems (Pallotti, 2015) and the morphological computations that need to be performed under communicative pressure (Lardiere, 2016). Hawkins and Casillas’s (2008) Contextual Complexity Hypothesis predicts that the probability with which an inflectional morpheme will be omitted by early-stage L2 English learners is a function of the number of contextual dependencies to be calculated. For example, there are more steps in the computation of the English 3<sup>rd</sup> person singular present subject-verb */-s/* than in the computation of *-ing* or *-Ø* (a bare verb), which may explain why learners are more likely to omit *-s* than to omit *-ing*. Complexity can also be related to the syntactic and discourse distribution of morphemes. Laleko and Polinsky (2016) found that L2 learners and heritage speakers of Korean and Japanese found morphological case markers that involve semantic and discourse computation (marked with topic) more difficult than case markers governed by syntactic constraints (nominative case).

Functional and cognitive approaches focus on the critical importance of processing for computing the mappings between concepts and morphemes from the input. Assuming an emergentist perspective, O’Grady (2008) and O’Grady et al. (2011) maintain that the language processing system has a key role in establishing what is initially acquired, what is subsequently retained or lost, and what is never acquired in the first place when considering how salience, frequency and transparency facilitate the establishment and strengthening of form-meaning mappings at the word and morpheme levels. Salience refers to acoustic prominence (*-ing* is more prominent and audible than *-s*); transparency to the one-to-one relationship between form and meaning, regularity to consistency and predictability in allomorphy involved in paradigms, and frequency to number of instances (types and tokens) in the input (see also Goldschneider and DeKeyser, 2001). The phenomena that are most susceptible to partial acquisition in heritage languages are those for which the form-meaning mappings are difficult to establish, either because the acoustic salience of a morpheme is weak (O’Grady et al., 2011), or because the precise semantic function or syntactic distribution may be difficult to figure out (Chung, 2016). These mappings require high frequency instantiations in the input, a condition that is not often met in heritage language and L2 learning. Longitudinal studies of child heritage speakers have found that when frequency and amount of input in the heritage language decrease with the shift to the majority language, inflectional morphology is very vulnerable at a young age (Silva-Corvalán, 2014), and instability persists into adulthood (Silva-Corvalán, 1991). In sum, both formal and functionalist approaches account for morphological errors and identify potential sources of difficulty related to the complexity

of form-meaning mappings, entrenched knowledge of another language, and processing computations.

Another common explanation for persistent morphological variability in L2 acquisition concerns age effects. Johnson and Newport (1989) argued that the non-native acquisition of morphology in L2 learners was related to biologically-determined maturational effects on input processing mechanisms. Children are eventually better at mastering morphology because of their limited processing abilities compared to adults (the Less is More Hypothesis). Since then, a prevalent stance has been that L2 learners' inability to (1) acquire features and abstract grammatical categories not instantiated in the L1 (Tsimpli and Dimitrakopoulou, 2007; Hawkins and Casillas, 2008), (2) achieve integrated knowledge of morphology (Jiang, 2004), and (3) process morphology like native speakers (Silva and Clahsen, 2008) is related to AoA after puberty (DeKeyser and Larson-Hall, 2005; Abrahamsson and Hyltenstam, 2009; Granena and Long, 2013). Others contend that ultimate attainment in L2 acquisition, and inability to reach native norms in all linguistic domains, is more readily explained by input and experience (Bialystok, 1997). Age effects are also relevant to explain the loss and weakening of their L1 in heritage speakers (Montrul, 2008): the younger the AoA of the majority language, the more likely the non-native acquisition of the heritage language (Yeni-Komshian et al., 2000), although input and experience also play a role (Jia and Aaronson, 2003). Heritage speakers contribute a different and unique angle on age effects because they illuminate how, despite L1 exposure since birth, restricted input in later childhood and adolescence greatly impacts the ultimate attainment of the heritage language in adulthood, especially at the morphological level.

## CASE MARKING IN HINDI

Case marks thematic roles (agent, patient, goal) linked to syntactic positions (subject, object, indirect object), and there is cross-linguistic variation in case systems with respect to how different languages mark overt case. Some languages present a nominative-accusative pattern and others an ergative-absolutive pattern, as illustrated in **Table 1**.

Nominative-accusative languages (e.g., Spanish, English, Russian, Greek, and German) generally mark subjects of transitive and intransitive verbs with nominative case, and objects with accusative case. Ergative-absolutive languages (e.g., Inuittitut, Dyirbal, and Basque) mark subjects of transitive verbs

with ergative case. Subjects of intransitive predicates and objects of transitive predicates are marked with absolutive case (Butt, 2006). Very few studies have investigated the acquisition of languages with ergative-absolutive patterns (Bavin and Stoll, 2013) and the present study provides new empirical evidence from Hindi. As a split ergative language, Hindi behaves morphologically as an ergative language in certain contexts and as a nominative-accusative language in others (Dixon, 1994; Bittner and Hale, 1996). The morphological and syntactic status of the ergative case (as structural or inherent) in Hindi and in other languages compared to other cases continues to be a topic of lively theoretical debate in various frameworks (Marantz, 1991; Davison, 2004; Anand and Nevins, 2006; Butt, 2006; Woolford, 2006; Keine, 2007; Coon, 2013). Our study is strictly concerned with the morphological expression of these cases and less with their syntactic status (structural, non-structural), or the consequences of one particular syntactic analysis over another. Specifically, we focus on the acquisition of the syntactic and semantic conditions for the morphological expression of the particle *-ne*, which marks ergative subject, and the particle *-ko*, appearing with all indirect objects, some subjects, and some direct objects. Therefore, we adopt a morphological account.

The ergative split in Hindi is conditioned by perfectivity. Ergative marking can only appear on subjects of transitive-perfective verbs as in (1), which shows the ergative particle *-ne* on the subject *Nikhil*. In addition, the object can be absolutive (i.e., no overt case and controlling verbal agreement). In perfective clauses, ergative subjects tend to be interpreted as agentive, or to have volitional control (In all other cases, the subject is zero marked, i.e., nominative). Example (2) has a verb in the imperfective, and the subject carries nominative case because imperfective predicates cannot license ergative marking on the subject. Example (3) is ungrammatical because the verb is intransitive (and perfective), and intransitive verbs do not license ergative marking *-ne* on the subject.

- (1) Nikhil-*ne*      akhbaar                      paRh-ii      hai.  
Nikhil.           newspaper.              read-Perf.      be.Pres.  
MSg-Erg      FSg.Nom                      FSg              3Sg  
'Nikhil has read the newspaper.'
- (2) Nikhil           akhbaar                      paRh-taa      hai.  
Nikhil.           newspaper.              read-Impf.      be.Pres.  
MSg.Nom      FSg.Nom                      MSg              3Sg  
'Nikhil reads the newspaper.'
- (3) \*Bela-*ne*      ghaNTaN                      dauR-ii/-aa.  
Bela.           for.                      run-Perf.FSg/Perf.MSg  
FSg-Erg      hours  
'Bela ran for hours.'

Case marking in Hindi interacts with verbal agreement. Main verbs and auxiliaries can agree with the subject (S-V agreement), with the object (O-V agreement), or with neither (default agreement). The verb agrees with nominative subjects. When subjects are ergative or dative the verb agrees with the nominative/absolutive object. If the subject and the object are overtly marked with ergative or accusative or dative, the verb

**TABLE 1 |** Case systems.

Nominative-accusative system		Ergative absolutive system	
Nominative case	Accusative case	Ergative case	Absolutive case
A	O	A	O
S			S

A = Subject of transitive verb, S = Subject of intransitive verb, O = Object (of transitive verb).



shows default, masculine singular agreement. Object agreement marks number and gender, while subject agreement marks number, gender, and person. We do not discuss agreement in this study, but see Montrul et al. (2012).

The other particle we investigate, *-ko*, appears with some direct objects, with indirect objects and with dative subjects. With direct objects, Hindi shows differential object marking (or DOM) (Aissen, 2003; de Swart and de Hoop, 2007), which is triggered by animacy (+ human) and specificity. Human, definite and specific direct objects must be overtly marked with *-ko*, as in (4), and are ungrammatical without *-ko* marking (With non-human animates, *-ko* is optional depending on the animal).

- (4) AaSaa-ne Niiraj-ko rokaa.  
Asha-Erg Niraj-DOM stopped  
'Asha stopped Niraj.'  
(cf. \*AaSaa-ne Niraj rokaa)

Human indefinite specific objects can be optionally *-ko* marked, as in (5).

- (5) Sudhaa ke gharwaaloN-ne us-ke liye laRke  
Sudha of relatives.MPI-Erg her-for boys.MPI  
dekhe(dekhaa)  
saw.MPI(saw.MSg)  
'Sudha's relatives saw boys for her (for marriage).'  
(laRkoN-ko)  
(boys.MPIobl-DOM)

With inanimate objects, *-ko* signals specificity. Inanimate, direct objects can be optionally marked with *-ko*, as in (6). If the object is *-ko*-marked it is interpreted as definite or specific; if it is unmarked it is non-specific.

- (6) Aashaa-ne rikshaa(/rikshe ko) rokaa  
Asha-Erg rikshaw (rikshaw DOM) stopped  
'Asha stopped a/the rickshaw.'

In general, non-specific or indefinite inanimate objects are mostly unacceptable with *-ko* marking, as in (7), but the acceptability of *-ko* in these cases has to be evaluated in context.

- (7) \*Sudhaa-ne ek caTTaan-ko dekhaa.  
Sudha-Erg a rock-DOM saw  
'Sudha saw a (non-specific) rock'

In sum, following Mohanan (1993), Aissen (2003), de Swart and de Hoop (2007), Dayal (2011), López (2012) and the judgments of the Hindi-speaking authors of this study, we assume these generalizations for *-ko* marking:

- (1) *-ko* is obligatory with personal pronouns and human proper names (and optional with non-human animates).
- (2) Definite, specific human DPs in object position take *-ko*.
- (3) *-ko* is optional elsewhere (but depends on context)

*-ko* is the overt morphological expression of accusative case and carries the feature [+ specific] when it attaches to (inanimate) direct objects. Hindi DOM is primarily triggered by animacy (+ human), which is an inherent feature of nouns, and also by specificity (de Swart and de Hoop, 2007), which is determined contextually.

Additionally, dative *-ko* can mark indirect objects (goals, beneficiaries), as in (8), and dative subjects (experiencers), as in (9). The marking of indirect objects and dative experiencers with *-ko* is obligatory, irrespective of animacy, definiteness or specificity. **Table 2** summarizes the distribution of the Hindi case particles discussed.

- (8) Manu-*ne* Niluu *ko* ticket dii.  
Manu-Erg Nilu-Dat ticket gave  
'Manu gave a ticket to Nilu.'  
(cf. \*Manu-*ne* Niluu ticket dii)
- (9) Manu-*ko* vah film pasand hai  
Manu-Dat that movie likes  
'Manu likes that movie.'  
(cf. \*Manu vah film pasand hai)

Keine (2007) presents a morphological analysis of Hindi split ergativity and the particles *-ne* and *-ko* within Distributed Morphology (Halle and Marantz, 1993) to account for why these markers sometimes are realized as zero. In Distributed Morphology, there is separation of syntax and morphology such that morphology is added post-syntactically. Affixes carry abstract syntactic and semantic features that can be more or less specified, and compete for lexical insertion post-syntactically depending on how well they match the formal semantic and syntactic features of the stem.

In Keine's analysis, which we assume for our study, ergative-*ne* is syntactically licensed in T if the predicate is perfective, as described in **Table 2**. Accusative *-ko* is licensed in V of transitive predicates; dative *-ko* of indirect objects in V of ditransitive predicates. Morphologically, both *-ne* and *-ko* alternate with the null marker, and Keine captures these patterns of overt/non-overt case alternations by means of morphological "impoverishment rules." In general, the overt markers are chosen in Hindi, but in certain contexts features are deleted, only allowing for the null or zero marker to be attached. For example, the impoverishment rule for ergative in (10) states that if the subject of a transitive verb is in an imperfective clause, then the case realization is zero instead of *-ne*, as in example (2):

- (10) [+ subject] → ø/[-perfective]

The impoverishment rule for accusative *-ko* states that if a direct object is inanimate and non-specific, the morphological case is realized as zero instead of *-ko*, as in example (2).

- (11) [+ oblique] → ø/[-human, -specific, + α]

The contextual features of these impoverishment rules capture the principles underlying the alternations between overt and zero markers, therefore giving rise to split ergativity as a morphological phenomenon. Keine (2007) does not discuss the *-ko* of dative experiencer subjects, other than saying that it is a special case because this type of *-ko* is lexically determined and does not alternate with zero. We assume that dative subject *-ko* is licensed lexically by the experiencer feature of the subject and by the lexical V (Davison, 2004).



**TABLE 2 |** Summary of the Hindi case particles examined in this study.

Particle	-ne	-ko <sub>1</sub>	-ko <sub>2</sub>	-ko <sub>3</sub>
m-case	Ergative	Accusative	Dative	Dative
Grammatical relation	Subject	Direct object	Indirect object	Subject
Thematic role	Agent	Patient/theme	Goal/beneficiary	Experiencer
Syntactic features	+ Subject + Oblique	– Subject + Oblique + $\alpha$	– Subject + Oblique – $\alpha$	+ Subject + Oblique
Semantic features	+ Perfective + volitionality	+ Human + Specific		– Perfective
Case valued in <sup>1</sup>	Perfective T	Transitive V	Ditransitive V	Lexical V

There are other case markers, such as instrumental -se, genitive -k and locative meN/par. <sup>1</sup> This is according to Keine (2007).

To summarize, the acquisition of ergativity raises questions about how monolingual and bilinguals identify the alignment patterns from the input and mark them correctly with ergative, nominative or absolutive case morphology. Not only does Hindi present both ergative-absolutive and nominative-accusative case patterns, but some cases can have multiple morphological realizations (overt/zero) and form-meaning mappings (one-to-one, as with *-ne*, and one-to-many, as with *-ko*). Case markers may not always be easy to perceive in the input: they vary in syntactic and semantic distribution and in frequency. Learners must implicitly perform distributional analyses of the input to figure out the structural differences between *-ko* as a marker of animacy with human direct objects and specificity with inanimate ones (DOM), but as an obligatory dative marker of *all* indirect objects and dative experiencers. This apparent variability in the Hindi system certainly presents a learning challenge for the acquisition and maintenance of case marking. As stated earlier, morphemes that are more frequent, map consistently to one meaning, and do not alternate with zero, are easier to acquire than morphemes that are less frequent in the input, map to more than one meaning, and alternate with zero (Kempe and MacWhinney, 1998; Goldschneider and DeKeyser, 2001; MacWhinney, 2008; O'Grady et al., 2011). On a formal account, morphemes that bundle more semantic and syntactic features and require multiple morphological computations are more likely to be omitted than morphemes requiring less steps in the computations (Hawkins and Casillas, 2008; Lardiere, 2016).

Paradoxically, languages that have rich morphology are easier to acquire than languages with sparser morphology, because the input provides many cues for morphological acquisition (Yang, 2002). But when input to the language is more restricted in terms of overall quantity and frequency, it will affect the abundance of morphological cues available and thus the degree of acquisition of morphology. Empirical evidence suggests that child learners of Hindi in India acquire and master ergative case marking relatively early while L2 learners and heritage speakers whose dominant language is English exhibit difficulty with accusative and ergative marking. Narasimhan (2005) studied three Hindi children in New Delhi, ages 1;7–3;9. The children made some omission errors with *-ne* but did not overgeneralize *-ne* to other predicates, and by the end of the observation period they showed between 80 and 100% accuracy on ergative marking. Under Keine's (2007) analysis, the children have correctly learned the impoverishment rule of *-ne* deletion with imperfective predicates.

Hindi is a null subject language and subjects are frequently dropped, as Narasimhan (2013) confirmed in the adult speech these children received, so the fact that Hindi children are so accurate at such an early age suggests that Hindi children are sensitive to perfective marking on the verb as a cue to ergativity. With respect to Hindi as a heritage language, Montrul et al. (2012) examined knowledge of case and agreement in oral narratives and grammaticality judgments. The Hindi heritage speakers produced and accepted ungrammatical sentences with omission of *-ne* and *-ko* with human specific direct objects, while the baseline Hindi-speaking adult immigrant group hardly omitted case markers in production or accepted ungrammatical sentences with omission in the judgment task. If overt *-ko* and ergative *-ne* marking is the default in these cases, following Keine's analysis, perhaps influence of English, which does not have overt case marking, is what underlies the high incidence of zero marking in bilinguals whose Hindi is the weaker language. Two early studies on the L2 acquisition of Hindi by English-speaking learners (Hansen, 1986; Lakshmanan, 1999) reported comprehension errors with subject and object relative clauses because the learners ignored (did not process) case marking (accusative *-ko*). Baten and Verbeke (2015) and Baten et al. (2016) found that Dutch L2 learners of Hindi have difficulty with ergative and DOM marking in oral production. While there is some independent evidence that the acquisition of Hindi morphology is problematic for both heritage speakers and L2 learners of Hindi, no study has directly compared the nature of L2 learners and heritage speakers' difficulty using the same methodology.

## THE STUDY

Our study investigates knowledge of morphological ergativity in Hindi (accuracy on *-ne*) and *-ko* marking with different NPs (dative subjects, direct object, and indirect objects) in English–Hindi bilinguals with different dominance patterns and levels of proficiency in Hindi, guided by the following research questions and hypotheses:

- (1) Does pattern of bilingual balance and proficiency in Hindi relate to knowledge of case marking in Hindi–English bilinguals?
- (2) In bilinguals with Hindi as the weaker language, does their knowledge of case marking differ as a function of AoA of Hindi (early in heritage speakers, late in L2 learners)?

Balanced bilinguals with higher proficiency in Hindi are expected to have more native-like knowledge of case marking in Hindi than bilinguals for whom Hindi is the weaker language. As for differences between L2 learners and heritage speakers, assuming that case marking in Hindi is acquired by age 3 (Narasimhan, 2005, 2013), if heritage speakers of Hindi acquired case morphology early in a naturalistic setting and *retained it*, they may have an advantage in overall accuracy over L2 learners with acquisition of Hindi in adulthood, even if input and use of Hindi decreased as the bilinguals got older. But if limited input and use of Hindi beyond early childhood contributed to heritage speakers not fully learning or forgetting the syntactic and semantic features of morphemes, no advantages over L2 learners are expected.

Given the specific complexity of the Hindi case system with respect to English, the dominant language, two other questions we examine are as follows:

- (3) If the unbalanced bilinguals make morphological errors in Hindi, will there be more omission or overgeneralization errors of the *-ne* and *-ko* particles?
- (4) Since *-ko* has multiple functions in Hindi, is the morphological realizations of *-ko* marking with direct object, indirect objects and dative experiencers more difficult to master in some syntactic contexts than in others?

Lardiere's (2009) Feature Reassembly Hypothesis is about linguistic representations whereas O'Grady et al.'s (2011) emergentist approach prioritizes the role of input and processing. These two models emphasize different aspects of the learning problem: what needs to be acquired (features and morpholexical forms) and on the basis of how it is acquired (noticing cues in the input). Both proposals make similar predictions regarding difficulties with different morphological markers, but for different reasons. We assume that ergative *-ne* and dative-experiencer *-ko* ( $ko_3$  in Table 2) are linked to agents of perfective predicates (ergative subjects) and experiencers of psychological predicates with stative verbs (dative subjects), respectively. Accusative *-ko* ( $ko_1$  in Table 2) is subject to NP constraints on definiteness and specificity with human and inanimate objects. Under certain semantic conditions ergative *-ne* and accusative *-ko* "appear" optional (realized as zero) when impoverishment rules apply (Keine, 2007). We further assume that *-ko* marking with indirect objects ( $-ko_2$  in Table 2) encodes syntactic features but no additional semantic features, being less structurally complex than accusative  $ko_1$ . Because it is consistently expressed as *-ko* (never zero) and most often refers to human goals/recipients, dative *-ko* with indirect objects is more reliable for learning than accusative  $-ko_1$ .

For Keine's (2007) analysis, *ne*-marking is the default for subjects and *-ko* marking is the default for human, specific direct objects: impoverishment rules in (10) and (11) apply with intransitive and imperfective predicates (leading to split ergativity) and when direct objects are inanimate and/or non-specific (DOM). If errors are observed, there will be overgeneralization (rather than omission) of *-ne* to intransitive

imperfective predicates and of *-ko* to inanimate and non-specific objects, respectively. But since the participants are bilingual and their stronger language is English, which does not have ergative case and DOM, dominant language transfer in this case may lead to significant more omission of the markers rather than to overgeneralization.

With respect to the specific complexity of the markers, we hypothesize that if L2 learners and heritage speakers make errors, these will be determined by the syntactic and semantic complexity of the markers. According to the Feature Reassembly Hypothesis, markers that bundle more semantic features will be more difficult to master than markers that bundle fewer features or only one. We thus expect higher accuracy with the *-ko* of indirect objects ( $ko_2$ ) than with ergative *-ne*, the *-ko* of dative experiencers ( $ko_3$ ) and the *-ko* of specific direct objects ( $ko_1$ ). From an input-based perspective,  $-ko_2$  with indirect objects is a more reliable cue than  $-ko_1$  with direct objects because all indirect objects are marked with *-ko* (i.e., it is not subject to any impoverished rule) whereas *some* direct objects are marked with zero. Therefore, this theoretical position also predicts higher accuracy on *-ko* marking of indirect objects than on the other three markers. Dominant language transfer can also account for omission of *-ko* with dative subjects, since English has nominative subjects with psych verbs. *-ko* marking with indirect objects is again predicted to be the easiest to be acquired because it is marked by a preposition (*to*) in NP PP configurations in English.

## Participants

Bilingual dominance in this study was determined by the linguistic and biographical characteristics of the bilinguals recruited, including AoA of the languages, place of upbringing, place of current residence, as well as specific linguistic measures of proficiency in Hindi. A total of 73 young adult Hindi-English bilinguals participated in this study. Based on language learning experience, place of current residence (United States vs. India) and self ratings on Hindi and English, the participants were grouped in three groups: a balanced bilingual group tested in India (the baseline group) ( $n = 23$ ) and two groups of English-Hindi bilinguals dominant in English (26 Hindi heritage speakers, 24 L2 learners of Hindi). All participants completed an extensive language background questionnaire, a written interview protocol that elicits short answer questions about demographic and biographical information, including information about the languages spoken at home, the activities performed in each language, the participants' current and past exposure to and use of Hindi and English at home and in other contexts (including school, travel), regular presence of grandparents, presence of older siblings, the languages used by parents with the heritage speakers at different times in childhood and the languages used by the heritage speakers with the parents. The questionnaire contains questions with Likert-scales eliciting information about perceived abilities in the two languages by skill (speaking, listening, reading, and writing), and estimates of quantity of input from estimates of the percentage of each language addressed to the participant, and the estimated amount of Hindi used by the participant beyond the home (TV, reading, internet, extracurricular activities, church), and with different

interlocutors (parents, siblings, grandparents, friends, other) who speak Hindi. Many of the questions in this questionnaire were not relevant for the L2 learners of Hindi, who were all born and raised in English-speaking homes in the United States.

The heritage speakers of Hindi, mean age 21.5 (range: 18–25), were recruited in Illinois and in New Jersey. They were simultaneous bilinguals exposed to English and Hindi in early childhood, born in the United States to highly educated Hindi-speaking parents (both father and mother). Due to the multilingual situation in India, the parents spoke both English and Hindi in addition to a regional South-Asian language (Punjabi, Gujarati, Marathi, Telugu, Tamil, among others). Some of these languages are ergative and others are not, but as we will see in the results, knowledge of these languages did not lead to variability with ergative marking in this group. Most of the heritage speakers ( $n = 20$ ) spoke English and Hindi before age 5 and the rest ( $n = 6$ ) spoke only Hindi. All Hindi heritage speakers were schooled in English and 18 indicated that they received from 2 to 10 h of instruction per week in Hindi as a heritage or foreign language in elementary and middle school through their parents. Use of Hindi during their lifetime was mostly with the parents and to a more limited extent with siblings. At present, 13 preferred to use English exclusively, while the rest would use more English than Hindi, depending on the situation. The heritage speakers were not taking Hindi classes at the time of testing, but they had all traveled to India at least once. When asked how they felt about Hindi, 4 (18%) indicated it was their native language and 22 (82%) their second language. Their mean self-assessments indicates that the majority of individuals in this group perceived Hindi as their less dominant language, and their impressions is corroborated by the biographical and language use information collected with the questionnaire.

The L2 learners of Hindi were graduate and undergraduate students ages 21 to 37 (mean: 26.24) taking Hindi as a foreign language in Illinois. Their mean length of exposure to Hindi was 4.2 years (range 1 to 7). They were all native speakers of English and started learning Hindi between the ages of 18 and 29 (mean age: 22). The learners were enrolled in advanced classes three or four times a week, which focused on reading, writing and speaking skills through culture. Eleven had traveled to India for 2 weeks to 5 months. Reasons for studying Hindi ranged from professional and academic (46%) to personal fulfillment (54%).

A main issue when doing experimental studies with heritage speakers is the baseline (Montrul, 2016b), and this depends on the objective of the study. Because our goal was not to examine the intergenerational language transmission in immigrants (for a study of intergenerational transmission see Montrul et al., 2015), we did not use a group of first generation adult immigrants in this study, and we chose to compare instead the heritage speakers and the L2 learners to age and SES matched peers in India, who are also bilingual in English and Hindi. The Hindi speakers from India were young university-educated adults between the ages of 18 and 25 residing in Delhi. They were fluent bilinguals in Hindi and English and, like the parents of the heritage speakers, most of them also spoke another South Asian language (Punjabi, Gujarati, Marathi, Tamil and Telugu,

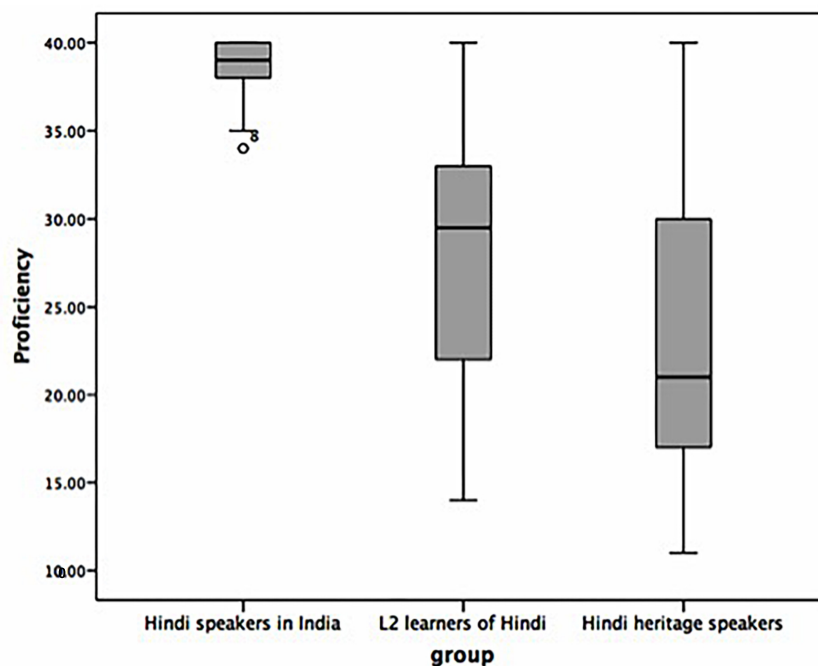
among others). It was not possible to control for what type of other South Asian languages the speakers knew. Sixteen (70%) reported that Hindi was their native language and 7 (30%) their second language. As far as patterns of language use at time of testing, 7 (30%) used Hindi the most in every day life, 4 (17%) used more English than Hindi, and the rest (52%) used the two languages on a daily basis.

The background questionnaire included self-rating scales on Hindi and English. Participants rated on a scale from 1 (none) to 5 (native ability) their overall perceived ability in English and in Hindi, in receptive (listening, reading) and productive (speaking, writing) skills. They also completed a written proficiency test, consisting of a cloze passage in Hindi with 40 blanks every seven words and three multiple-choice responses per blank (same cloze test was used in Montrul et al., 2012, 2015). This cloze task was created by one of the authors of this study and was piloted with native and non-native speakers of Hindi. Reliability statistics (Cronbach alpha) run on the responses of the cloze test yielded a coefficient above 0.80. Although we did collect information about the frequency of use of Hindi and English for all the speakers tested, following Montrul (2016a) we assessed dominance quantitatively, by combining the scores from the self-ratings and accuracy in the Hindi written proficiency measure. Reported amount of input and use of the language, an important dimension of dominance (Montrul, 2016a), largely corroborated the self-ratings and general Hindi proficiency scores.

**Table 3** presents the self-ratings in each language for the three groups. **Figure 1** shows the dominance patterns of the three groups based on the overall self-ratings. Comparison of mean self-ratings in English and in Hindi showed that the Hindi speakers in India self-evaluated their overall Hindi proficiency as high as their English [paired samples  $t$ -test:  $t(22) = 1.19$ ,  $p > 0.05$ ], and evaluated similarly their four skills in each language (all  $ps > 0.05$ ). Therefore, they are considered balanced in English and Hindi bilinguals for this study. The heritage speakers and the L2 learners self-rated their English at native level and their Hindi significantly lower [paired samples  $t$ -tests: L2 learners  $t(23) = 13.515$ ,  $p < 0.0001$ , heritage speakers  $t(25) = 8.17$ ,  $p < 0.0001$ ]. Thus, they considered unbalanced bilinguals for this study, with English as dominant language and Hindi as their weaker language. The L2 learners and the heritage speakers assigned lower ratings to their Hindi than the speakers in India (balanced bilinguals) (one way ANOVAs and Tukey *post hoc* tests, all  $ps < 0.0001$ ). Except for speaking, which the L2 learners and the heritage speakers rated similarly (2.88 and 2.58,  $p > 0.05$ ), the two groups differed on their assessments of reading, listening and writing skills (all  $ps < 0.05$ ). The heritage speakers rated their listening skills higher (3.46) than the L2 learners (2.48), whereas the L2 learners rated their reading and writing skills (3 and 3.04) higher than the heritage speakers (1.88 and 1.65). The heritage speakers attended English only schools but many said they received instruction in Hindi during the elementary and middle school period through their parents, whereas the L2 learners learned to read and write the Hindi script in the classroom. This difference among skills within and between the two groups confirms the

**TABLE 3 |** Mean self-ratings in Hindi and English language skills (1 = none-limited ability, 5 = native ability), *SDs* are in parentheses.

Language	Skill	Groups					
		Hindi speakers in India ( <i>n</i> = 23)		Hindi heritage speakers ( <i>n</i> = 26)		L2 learners of Hindi ( <i>n</i> = 24)	
		Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
English	Reading	4.26	(1.09)	5	(0)	5	(0)
	Speaking	4	(1.04)	5	(0)	5	(0)
	Listening	4.30	(1.06)	5	(0)	5	(0)
	Writing	4.04	(1.06)	4.92	(0.27)	5	(0)
	Overall	4.34	(1.07)	5	(0)	5	(0)
Hindi	Reading	4	(1.12)	1.88	(0.95)	3	(0.83)
	Speaking	4.13	(0.96)	2.88	(1.21)	2.58	(0.97)
	Listening	4.43	(0.78)	3.46	(1.17)	2.58	(0.82)
	Writing	3.69	(1.25)	1.65	(0.79)	3.04	(0.80)
	Overall	4.65	(0.57)	3.26	(1.07)	3.12	(0.67)

**FIGURE 2 |** Mean distribution of scores in the written Hindi proficiency cloze test.

common profile of L2 learners and many heritage speakers in their weaker language.

The results of the Hindi proficiency test in **Figure 2** reflected similar differences between the three groups [one way ANOVA,  $F(2,72) = 32.08$ ,  $p < 0.0001$ ]. From a total maximum of 40 points, the mean score for the speakers in India was 38.56 (34–40, *SD*: 1.82), 24.11 (11–40, *SD*: 8.60) for the heritage speakers, and 27.91 (14–40, *SD*: 6.62) for the L2 learners. Multiple comparisons showed no statistical difference between the proficiency scores of the L2 learners and the heritage speakers ( $p > 0.05$ ). The overall Hindi proficiency ratings and the scores

on the written proficiency test correlated positively,  $r$  (two-tailed) for the L2 learners = 0.59,  $p = 0.001$  and for the heritage speakers = 0.55,  $p = 0.006$ .

## Tasks

An oral production task and a bimodal AJT were used to assess knowledge of morphological case. The oral production task elicited differential object marking and dative case, and consisted of pictures with two participants and transitive verbs requiring animate (human) or inanimate objects and verbs that take dative subjects. The task included 35



sentences: 7 with dative subjects, 14 with human objects, 14 with inanimate objects. Participants were asked to describe the pictures using the past tense, with many opportunities to use perfective predicates. Therefore, we examined the production of ergative marking in the same task. Responses were audio-recorded, transcribed, and analyzed for correct suppliance, omission, or overgeneralization of the case markers *-ko* and *-ne*.

The AJT (Montrul et al., 2012) complemented the results obtained with elicited production. The AJT was bimodal, with stimulus presentation in visual and auditory modality. Since L2 learners tend to do better in written than in auditory tasks whereas heritage speakers do better in auditory than in written tasks (Montrul et al., 2008), the bimodal presentation was done to not advantage or disadvantage the unbalanced bilingual groups with respect to each other. There were 216 sentences (half target, half fillers, half grammatical, half ungrammatical/infelicitous) divided into 24 types, with 6–12 token sentences per type, depending on the structure. Sentence types included minimal pairs with correct and incorrect uses of ergative case and verb transitivity in simple and compound verbs, and sentence types with human animate and inanimate, specific/non-specific direct objects, indirect objects and dative subjects, where the presence and omission of the case marker *-ko* were manipulated, like the examples presented in (1) to (9). For the sentences testing ergativity, we only manipulated transitivity; we did not manipulate perfectivity due to the length of the test. However, perfectivity errors were evaluated in the oral task. Even though the stimuli consisted of minimal pairs (the same verbs and sentence structure with and without the relevant morphology), the sentences were presented in randomized order (not in pairs), and each sentence was judged independently. The AJT was administered through the web interface Survey Gizmo. Each sentence was presented in Hindi script and with an audio player below. Participants were instructed to read each sentence and play the sound file before rating each sentence on a 1–4 scale (1 = completely unacceptable to 4 = perfectly acceptable). The task was self-paced and did not measure reaction times (completion time was about 20–30 min). Participants could not go back and compare sentences: once a sentence was rated, it disappeared from the screen.

## RESULTS

### The Elicited Production Task

Although the initial pool of L2 learners was 24, we have results for 19 learners. Two audio files were corrupted, and three learners who were not confident in their oral skills refused to complete the task. Transcriptions were coded for ergative *-ne* and accusative/dative *-ko* marking in obligatory contexts and for potential omission errors, and in non-obligatory contexts for potential overgeneralization errors. For ergativity, we coded presence or absence of *-ne* and its accuracy (correct/incorrect) based on transitivity (transitive/intransitive) and perfectivity (perfective-non-perfective) of all verbs. For *-ko* we coded presence or absence of *-ko*, its accuracy

(correct/incorrect) based on type of NP (direct object, indirect object, dative subjects) and animacy of the direct objects (human animate, non-human animate, and inanimate). Transcriptions, coding and inter-rater reliability checks were done by two of the Hindi-speakers in our team. Since this was production, the number of relevant tokens produced would be different for each speaker.

Most of the participants responded in the past and used many instances of ergative case, but some did not use past or ergative marking. Therefore, the results are based on the number of participants in each group who produced ergative *-ne* marking in required transitive perfective contexts. The number of observations included in the analysis was 2210. There were very few instances of intransitive predicates with no ergative marking produced in the entire data, all correct (1 instance from 1 heritage speaker, 5 instances from 5 speakers from India and 3 instances by 3 L2 learners). We counted overgeneralization errors of *-ne* with sentences in non-perfective contexts (present, progressive, or imperfective). Several heritage speakers and L2 learners made omission errors with perfective predicates, as in (12) and (13), or overgeneralizations of *-ne* to imperfectives, as shown in (14) and (15).

- (12) \***sarah**- $\varphi$  eva ko khiiNc-aa thaa  
 Sarah Eva ACC pull-Perf.MSg Pst.MSg  
 ‘Sarah had pulled Eva.’  
 (Hindi heritage speaker)
- (13) \***bill**- $\varphi$  sara ko rulaa-yaa  
 Bill Sara ACC make.cry-Perf.MSg  
 ‘Bill made Sara cry.’  
 (L2 learners of Hindi)
- (14) tom \***ne** stephanie ko cup kar rahaa hai  
 Tom Erg Stephanie ACC pacify do Prog.MSg Pres.Sg  
 ‘Tom is pacifying Stephanie.’  
 (Hindi heritage speaker)
- (15) john \***ne** esha ko uThaa rahaa hai  
 John Erg Esha ACC pick Prog.MSg Pres.Sg  
 ‘John is picking Esha up.’  
 (L2 learner of Hindi)

The raw data were analyzed using binomial linear mixed-effects models (Jaeger, 2008) in R Core Team (2014) on categorical data (correct, incorrect), better suited to analyze categorical data and unbalanced data (Jaeger, 2008, p. 436). All independent variables were added to the model following a stepwise procedure, and subsequently, models containing interactions between factors were also incorporated to the analysis. In identifying the best-fitted model for our data, all nested models were compared using the function ANOVA. The most reliable model was chosen based on lowest AIC values.

The best model for ergativity marking included group (heritage speakers, L2 learners, speakers from India) and aspect (perfective, non-perfective) as fixed effects, and participants and items (random intercepts only) as intercepts. The dependent

**TABLE 4 |** Elicited Production Task.

Groups	N	Accuracy on <i>-ne</i> with transitive predicates	Total error rate	Omission of <i>-ne</i> with perfective predicates	Extension of <i>-ne</i> to imperfective predicates
Hindi speakers in India	23	99.1	0.9	0.02	0.00
Hindi heritage speakers	26	81.5	18.5	42.31	0.75
L2 learners of Hindi	19	76.5	23.5	19.41	62.5

Percentage Accuracy on ergative *-ne* marking.

variable was accuracy on *-ne* marking. **Table 4** shows accuracy on ergative *-ne* marking with transitive, perfective predicates and rates of omission, and overgeneralization of *-ne* to non-perfective predicates in oral production.

The percentage accuracy of *-ne* and the total error rate was significant by group ( $\beta = 3.25$ ,  $SE = 0.89$ ,  $z = 3.639$ ,  $p < 0.0001$ ). There was a main effect for aspect ( $\beta = 3.52$ ,  $SE = 0.37$ ,  $z = -9.50$ ,  $p < 0.0001$ ) and an aspect by group interaction ( $\beta = -7.79$ ,  $SE = 0.96$ ,  $z = -8.079$ ,  $p < 0.0001$ ). Tukey *post hoc* comparisons revealed that overall accuracy on *ne*-was significant different ( $p < 0.05$ ) between the three groups, but the interaction indicated that the heritage speakers omitted *-ne* (42.3%) with transitive perfective predicates more than the L2 learners (19.4%) ( $p < 0.0001$ ). The error types were examined by omissions and overgeneralizations. Although instances of transitive imperfective predicates with *-ne* were very few in the data (total 18), the L2 learners produced significantly more overgeneralization errors with *-ne* (62.5%) with imperfective predicates than the heritage speakers (0.75%) ( $p < 0.0001$ ). The heritage speakers showed the opposite pattern: when errors were made, these were more of omission than of overgeneralization.

Next, we analyzed the use of *-ko* marking with direct objects, which is obligatory if the object is human animate and specific. The number of observations included in the analysis was 1702. Most of the examples included names or referred to people (grandmother, hunter). Examples (16) and (17) are errors of omission of *-ko* with human objects.

- (16) \*Teacher- $\varphi$  khuS kar rahe haiN chaatra  
 teacher please do Prog.MPI Pres.MPI students  
 OBJ V SUBJ  
 ‘The students are making the teacher happy.’  
 (Hindi heritage speaker)

- (17) grandmother \*Albert laa-ii  
 grandmother Albert bring-Perf. FSg  
 SUBJ OBJ V  
 ‘The grandmother brought Albert.’  
 (L2 learners of Hindi)

We included sentences with inanimate objects but because Hindi does not have articles, the slides only listed the name of the object. Specific inanimate objects are marked with *-ko*; non-specific objects are unmarked in Hindi. So, if the participants chose to make the object specific, they would use *-ko* and if they made the object non-specific, they would not mark it with

*-ko*. Sentences were not presented in context, so the use of *-ko* with inanimate objects was optional, depending on whether the participant meant the object to be specific or not, as in (18) non-specific unmarked and (19) specific, marked.

- (18) Pati patnii form bhar-eNge  
 husband wife form fill-Fut.MPI  
 ‘The husband and wife will fill a form.’  
 (Hindi heritage speaker)
- (19) aadmii ne chaate ko uThaa-yaa  
 man Erg umbrella ACC pick-Perf.MSg  
 ‘The man picked up the umbrella.’  
 (L2 learner of Hindi)

Finally, *-ko* marking is also obligatory with dative subjects, but L2 learners and heritage speakers produced omission errors with these predicates, as in (20) and (21).

- (20) maaN kii madad se \*ye- $\varphi$  garv huaa  
 mother of help with he proud happen.Perf.MSg  
 ‘He became proud with mother’s help.’  
 (L2 learner of Hindi)
- (21) \*maa- $\varphi$  beTe par garv ho rahii hai  
 mother son at proud be Prog.FSg Pres.Sg  
 ‘The mother is feeling proud of her son.’  
 (Hindi heritage speaker)

**Table 5** shows the percentage production of *-ko* marking by NP type (direct objects, indirect objects and dative subjects) and **Table 6** shows production of *-ko* by the animacy of the direct object (Human, non-human animate, and inanimate).

We conducted two binomial mixed effects models with *ko*-accuracy as dependent variable. The first one included group and NP type as fixed effects, with participants and items as random effects. This model found a main effect for group ( $\beta = -1.91$ ,  $SE = 0.68$ ,  $z = 2.79$ ,  $p < 0.01$ ) and an NP type by group interaction ( $\beta = 2.20$ ,  $SE = 0.49$ ,  $z = 4.419$ ,  $p < 0.0001$ ). The main effect by group found that the L2 Hindi speakers were statistically significant from the L2 learners and the heritage speakers ( $p < 0.0001$ ). The groups by NP type interaction found that the heritage speakers omitted *-ko* with dative subjects more than with direct and indirect objects ( $p < 0.0001$ ). The second binomial fixed effects model included type of direct object and group as fixed effects with subject and items as random intercepts. Accuracy production of *-ko* was the dependent variable. This model found a significant main effect for animacy ( $\beta = 3.75$ ,

**TABLE 5 |** Elicited Production Task.

Groups	N	Direct object	Indirect object	Dative subject
Hindi speakers in India	23	100	100	98.96
Hindi heritage speakers	26	83.87	91.70	66.19
L2 learners of Hindi	19	88.23	93.55	86.04

Mean percentage accuracy accusative/dative -ko marking by NP type.

**TABLE 6 |** Elicited Production Task.

Groups	N	Human	Non-human animate	Inanimate
Hindi speakers in India	23	100	100	100
Hindi heritage speakers	26	85.00	79.16	98.36
L2 learners of Hindi	19	86.38	76.92	100

Mean percentage accuracy accusative-ko marking by animacy.

$SE = 0.48$ ,  $z = 7.67$ ,  $p < 0.0001$ ), no main effects for groups, and no interactions. Although the speakers from India performed at ceiling, there were no differences between the L2 learners and the heritage speakers in their overall accuracy: both groups omitted -ko with animate direct objects to the same extent (human: 16.13% heritage speakers, 11.77% L2 learners; and non-human: 20.84% heritage speakers, 23.08% L2 learners). There were very few overgeneralizations of -ko to inanimate objects. In general, the data in **Tables 5, 6** show that heritage speakers and L2 speakers, whose weaker language is Hindi, omitted obligatory accusative case marking with human direct objects and dative case with dative subjects, unlike the speakers from India (balanced bilinguals).

Summarizing, with respect to ergative marking, there were more omission than overgeneralization errors of -ne for the heritage speakers. With accusative -ko, the two unbalanced bilingual groups produced omission errors with human animate objects to the same extent (-ko with non-human objects is more variable). The heritage speakers produced a few -ko marking errors with inanimate objects compared to the L2 learners and the speakers from India, but these cannot necessarily be considered errors if the descriptions were meant to be specific, since the sentences were not ungrammatical. Finally, the L2 learners and the heritage speakers omitted -ko with dative subjects more than with indirect objects.

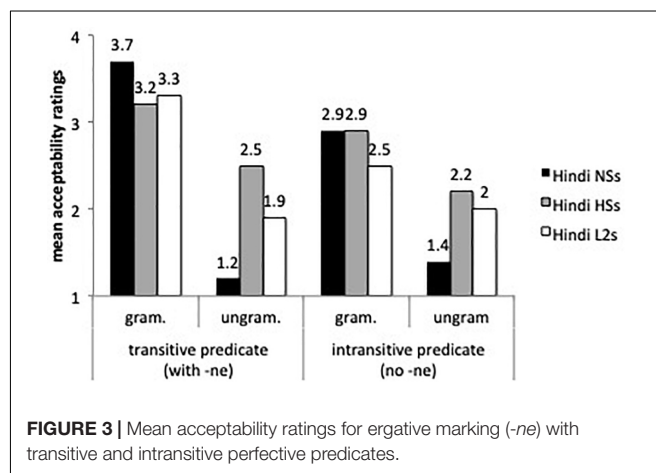
## The Acceptability Judgment Task (AJT)

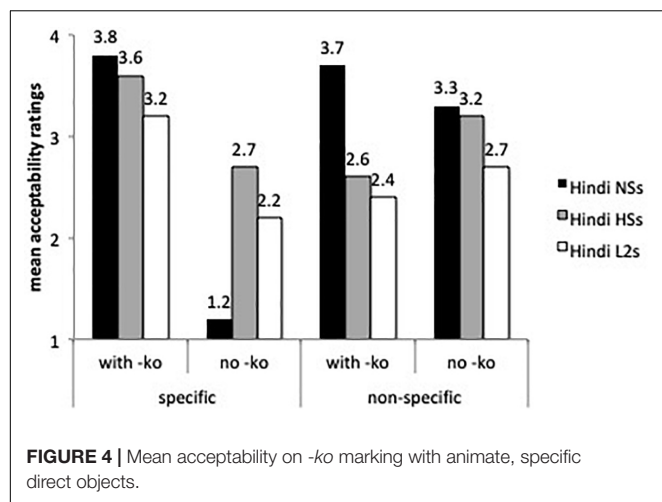
In the AJT, grammatical and ungrammatical sentences manipulating the markers -ne and -ko were judged on a scale from 1 to 4. Acceptability ratings were submitted to ordinal regression mixed effects models in R. One model included acceptability ratings as dependent variable, sentence type, grammaticality and group as fixed factors, and participants and items as random intercepts. The model found a main effect for group ( $\beta = 4.36$ ,  $SE = 1.78$ ,  $t = 2.45$ ,  $p < 0.01$ ), for sentences ( $\beta = -1.71$ ,  $SE = 4.40$ ,  $t = -3.887$ ,  $p < 0.0001$ ) and a group by sentences interaction ( $\beta = 7.15$ ,

$SE = 1.98$ ,  $t = 3.605$ ,  $p < 0.0001$ ). To investigate the group by sentences interaction further, we ran models on -ne and -ko sentences separately.

With respect to -ne marking (ergativity), we tested transitive and intransitive verbs, in both simple and compound verbs in the perfective form. **Figure 3** displays the mean acceptability ratings for transitive and intransitive predicates.

A mixed effects model with group by transitivity and grammaticality showed main effects for grammaticality ( $\beta = -1.083$ ,  $SE = 7.16$ ,  $t = -15.116$ ,  $p < 0.00001$ ), and group ( $\beta = 4.81$ ,  $SE = 1.05$ ,  $t = -0.81$ ,  $p < 0.0001$ ), and a group by grammaticality interaction ( $\beta = 3.354$ ,  $SE = 5.66$ ,  $t = 5.924$ ,  $p < 0.0001$ ). The heritage speakers and L2 learners' ratings were statistically different from those of the speakers in India ( $p < 0.001$ ), but Tukey *post hoc* tests revealed no differences between the L2 learners and the heritage speakers ( $ps > 0.05$ ). The group by grammaticality interaction indicated that the Hindi speakers from India, the heritage speakers and the L2 speakers differed in their ratings of ungrammatical sentences. The heritage speakers assigned higher acceptability ratings to ungrammatical sentences with transitive and intransitive predicates than the native speakers from India ( $\beta = -1.12$ ,  $SE = 5.95$ ,  $t = -18.932$ ,  $p < 0.0001$ ) and the L2 learners ( $\beta = 3.354$ ,  $SE = 5.66$ ,  $t = 5.924$ ,  $p < 0.0001$ ). As for omission of -ne with transitive perfective predicates, the results confirm the findings of the production task: the L2 learners and the heritage speakers assigned higher acceptability ratings to ungrammatical sentences with omission of -ne than the speakers from India, and the difference between the experimental groups' ratings was not significant (all  $ps > 0.05$ ). Heritage speakers and L2 learners were also more accepting of intransitive perfective predicates with -ne (i.e., overgeneralization errors) than the speakers from India ( $p = 0.01$ ), according to Tukey *post hoc* tests, suggesting unstable knowledge of -ne marking in intransitive predicates as well. For the heritage speakers and for the L2 learners we conducted pairwise comparisons of the two ungrammatical sentences: transitive predicates without -ne (omission) and intransitive predicates with -ne (overgeneralization) and there were no statistical differences between the ratings for either group.





**Figure 4** depicts the results of *-ko* marking with animate (human) specific direct objects, which is required in all these cases.

According to the best-fitted fixed effects model ran on animate objects (with group, sentences and grammaticality as fixed factors and subject and items as random intercepts), the speakers from India differed significantly from the heritage speakers and the L2 learners ( $\beta = 0.98$ ,  $SE = 0.19$ ,  $t = 5.15$ ,  $p < 0.0001$ ). Human specific objects with *-ko* received overall higher ratings than human non-specific objects with *-ko* ( $\beta = 0.861$ ,  $SE = 0.22$ ,  $t = 3.849$ ,  $p < 0.0001$ ). There was a sentence type by group interaction: the L2 learners assigned lower acceptability ratings to grammatical sentences with *-ko*-marked human non-specific direct objects than the Hindi speakers from India ( $\beta = 0.41$ ,  $SE = 0.20$ ,  $t = -2.057$ ,  $p < 0.01$ ). Even though the three groups rated grammatical and ungrammatical *-ko* with human specific objects differently (the comparison of grammatical and ungrammatical sentences was significant at the  $p < 0.0001$ ), the L2 learners and the heritage speakers were more accepting of *-ko* omission with human specific direct objects than the speakers from India ( $\beta = -1.356$ ,  $SE = 0.35$ ,  $t = -3.867$ ,  $p < 0.0001$ ), a result that confirms the omission errors found in the oral task. Other contrasts were not significant.

**Figure 4** displays ratings on inanimate objects, both specific and non-specific, with and without *-ko* marking. The mixed effects model found a main effect for group ( $\beta = 0.73$ ,  $SE = 0.16$ ,  $t = 4.346$ ,  $p < 0.0001$ ), for sentences ( $\beta = 0.5069$ ,  $SE = 0.22$ ,  $t = 2.216$ ,  $p < 0.01$ ), and a sentence by group interaction ( $\beta = 0.44$ ,  $SE = 0.17$ ,  $t = 2.49$ ,  $p < 0.01$ ). As the speakers from India's ratings show, *-ko* is more likely to be dropped with inanimate objects than with human specific direct objects (see **Figure 3**). The heritage speakers did not differ from the speakers from India with specific and non-specific inanimate objects, and the L2 learners were less accepting of *-ko* marking than the heritage speakers ( $\beta = -0.56$ ,  $SE = 0.22$ ,  $t = 2.48$ ,  $p < 0.01$ ) with specific inanimate objects. Although the speakers from India assigned higher ratings to inanimate non-specific objects with *-ko*

marking than expected, our task presented sentences in isolation, and without context these sentences can be assumed to involve specific objects and hence acceptable.

Recall that human, specific direct objects with no *-ko* are ungrammatical (omission error) (**Figure 4**) and in principle inanimate non-specific objects with *-ko* are ungrammatical (potential overgeneralization error) (**Figure 5**). The acceptability ratings on these sentences were not significant.

**Figure 6** shows the acceptability of *-ko* marking with indirect objects and dative subjects. The mixed effects model performed on these sentence types found a main effect for group ( $\beta = 0.52$ ,  $SE = 0.15$ ,  $t = 3.316$ ,  $p < 0.0001$ ), by sentence type ( $\beta = -1.77$ ,  $SE = 0.17$ ,  $t = -10.125$ ,  $p < 0.0001$ ) and a group by sentence type interaction. All three groups seem to know that *-ko* is grammatical with these two sentence types. The heritage speakers and the L2 learners were more accepting of indirect objects ( $\beta = -1.108$ ,  $SE = 0.19$ ,  $t = -5.778$ ,  $p < 0.0001$ ) and of dative subjects without *-ko* (i.e., omission) ( $\beta = -0.95$ ,  $SE = 0.18$ ,  $t = -5.158$ ,  $p < 0.0001$ ) than the speakers from India. The L2 learners and the heritage speakers did not differ from each other (Tukey tests non-significant).

Finally, **Figure 7** compares the mean acceptability ratings of ungrammatical sentences with case omission (errors) with ergative subjects (*-ne*), animate specific direct objects ( $ko_1$ ), indirect objects ( $ko_2$ ) and dative subjects ( $ko_3$ ) by the two experimental groups. Per our hypothesis we expected lower ratings (i.e., less acceptance of omission) with indirect objects ( $ko_2$ ) than with ergative *-ne*, dative experiencers ( $ko_3$ ) and specific direct objects ( $ko_1$ ).

A liner mixed effects model with just the L2 learners and the heritage speakers revealed no main effect for group but differences for sentence types. Indirect objects were rated differently from direct objects ( $\beta = 0.42$ ,  $SE = 0.16$ ,  $t = 2.619$ ,  $p < 0.05$ ) and dative subjects ( $\beta = 0.353$ ,  $SE = 0.16$ ,  $t = 2.154$ ,  $p < 0.05$ ), supporting our hypothesis. The L2 learners and the heritage speakers assigned higher acceptability ratings to ungrammatical sentences with omissions of ergative *-ne*, accusative *-ko\_1* and dative *-ko\_3* with dative experiencers than of dative *-ko\_2* with indirect objects (all comparisons significant at  $p < 0.05$ ), as predicted.

Summarizing the findings of the AJT, we found evidence of omission of *-ne* with transitive perfective predicates and overgeneralization errors involving ergative marking with intransitive perfective predicates, although acceptability ratings for omission and overgeneralization errors were not significant for the L2 learners and the heritage speakers. The L2 learners and heritage speakers accepted errors of omission of *-ko* with human specific direct objects. Potential errors of overgeneralization of *-ko* to inanimate non-specific contexts were harder to assess because the sentences were presented in isolation, and could still receive a specific reading. The heritage speakers were more native-like in their acceptability of sentences with *-ko* as a marker of specificity with inanimate objects than the L2 learners. As for case omissions by syntactic context, indirect objects received the lowest ratings, compared to the other three conditions: ergative *-ne*, *-ko* with direct objects and *-ko* with dative experiencers.



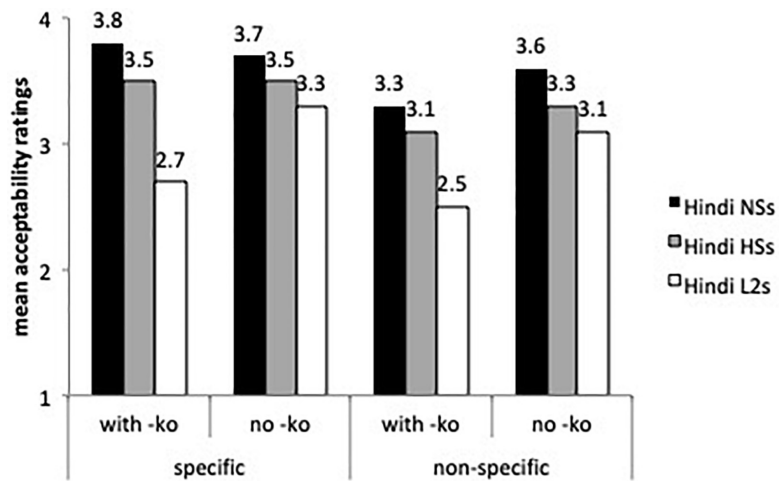


FIGURE 5 | Mean acceptability on *-ko* marking with inanimate objects.

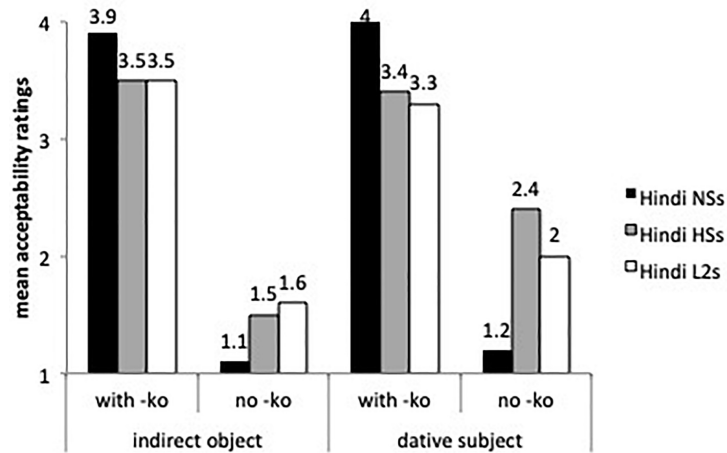


FIGURE 6 | Mean acceptability ratings on indirect objects and dative subjects with and without *-ko*.

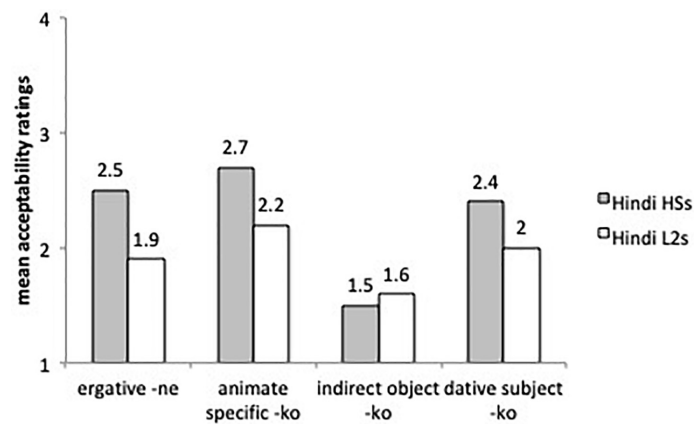


FIGURE 7 | Mean acceptability ratings on ungrammatical sentences with omission of obligatory case marking.

## DISCUSSION

The two main findings of our study were first, that morphological accuracy with Hindi case marking differed by group of bilinguals and second, that there were linguistic effects on the acquisition of the different case markers examined, modulated by semantic and syntactic complexity and input frequency.

The Hindi-English speakers from India outperformed the L2 learners of Hindi and the Hindi heritage speakers. The multilingual profile of the speakers from India is the typical reality of Hindi speakers in India, especially of the SES we tested. These speakers performed at ceiling, suggesting that knowledge of another South Asian language (some of which were not ergative) had no effect on their morphological accuracy in Hindi. By contrast, the heritage speakers and the L2 learners in our study made and accepted morphological case errors in Hindi. The three bilingual groups differed in several biographical variables, such as place of residence and upbringing, AoA of Hindi, and context of learning (naturalistic, instructed). If AoA of Hindi and context of learning were to explain the results, the speakers from India and the Hindi heritage speakers, both exposed to Hindi in a naturalistic setting early in life, would pattern together in their oral production and grammaticality judgments of Hindi case morphology. However, we found that the performance of L2 learners and the heritage speakers, who reside in the United States and were exposed to and currently use less Hindi than English and have lower proficiency in Hindi than English compared to the speakers from India (i.e., unbalanced bilinguals), was very similar. Thus, early AoA in a naturalistic setting did not matter for the acquisition of morphology. This result, to us, implies an overall dominance effect: the balanced bilinguals differed from the unbalanced bilinguals in their production and knowledge of the morphological complexity of the Hindi case system. This result is consistent with previous findings of early AoA effects for phonology but not for morphosyntactic knowledge in heritage speakers and L2 learners using production and off-line grammaticality judgment tasks (e.g., Au et al., 2002). O'Grady et al. (2001) also found no differences between L2 learners and heritage speakers of Korean on morphological case markers and relative clauses. At the same time, we note that the Hindi-dominant group and the heritage Hindi group did not only differ with regard to language dominance, but also with regard to the amount of exposure to Hindi they experience in their everyday lives. Thus, it is possible that the between-group difference that emerged in our study is actually not caused by language dominance alone, but is instead due to differences in the amount of exposure and current use of Hindi, which are experience-based components of dominance (Montrul, 2016a). Paradis (2010) study on language exposure, complexity and task type, suggests that exposure impacts the acquisition of morphology in school-age bilingual children. The children with highest exposure to English were also the most dominant in English and the ones that approached monolingual English norms more closely. Bedore et al. (2012) found that in pre-kindergarten and kindergarten-age children current

language use was a better predictor of language dominance with respect to morphosyntactic measures than age of first exposure. Unfortunately, we know of no studies that tease apart exposure and language use from dominance in adults, which would be useful to establish. Still, the Hindi-dominant group, which lived in India, used Hindi more often currently than the Hindi heritage speakers. Given this state of affairs, we thus consider it more likely for now that the differences in our study are due to overall language dominance.

Amount of exposure and use of Hindi may be the reason for the morphological inaccuracies found in the two unbalanced bilingual groups. The heritage speakers in this study were exposed to Hindi naturalistically since birth for almost 20 years, but they clearly do not master case, just like the L2 learners with less than 7 years of Hindi instruction. The non-target acquisition of case manifested by the heritage speakers can be explained by reduced exposure to and use of the family language during childhood in a language-minority situation. Montrul et al. (2012) showed that adult Hindi-speaking immigrants were native-like with all these case particles, suggesting that case marking is present in the input to Hindi speakers. Therefore, it is possible that being exposed to Hindi only through the parents and using it less frequently than English may have impacted the heritage speakers' opportunity to master case marking at native levels by adulthood. Since morphological learning depends on frequency and distribution in the input (Yang, 2002), an explanation of reduced input in childhood is compatible with the findings of the heritage speakers (O'Grady et al., 2011).

In L2 acquisition, on the other hand, non-native attainment may have two possible sources: limited exposure and restricted use of the L2 in an instructional setting (Bialystok and Hakuta, 1999), as well as the maloperation of the implicit language learning mechanisms available in childhood (DeKeyser, 2013). Our L2 learners were exposed to Hindi for 7 years at most, predominantly in an instructed setting a few hours a week. It is possible that with more input and use the L2 learners could reach the level of morphological accuracy of the Hindi speakers in India. Very advanced and near-native English-speaking L2 learners of Hindi would need to be tested to confirm this possibility. Since we did not use tests of implicit knowledge or online processing (e.g., timed grammaticality judgment tasks, self-paced reading tasks) we are unable to corroborate whether L2 learners and heritage speakers use different morphological processing mechanisms.

While finding that balanced bilinguals show better command of morphology than unbalanced bilinguals with less exposure and use of the language may be obvious, our study aimed to understand how linguistic factors may affect morphological acquisition in the weaker language. We hypothesized that the degree of accuracy on the case markers would vary as a function of syntactic and semantic complexity, and frequency in the input. The Feature Reassembly Hypothesis (Lardiere, 2009) and the Contextual Complexity Hypothesis (Hawkins and Casillas, 2008) predicted that markers that bundle more semantic and syntactic features will be more difficult to master

and require more morphological computations than markers bundling fewer features (see Hawkins and Casillas, 2008 for details). This is also determined by whether or not the features and feature bundles exist in the bilinguals' other language. Based on their different feature specifications, we expected higher accuracy with indirect objects ( $ko_2$ ) than with ergative *-ne*, dative experiencers ( $ko_3$ ) and specific direct objects ( $ko_1$ ). From an input-based perspective (O'Grady et al., 2011), *-ko* with indirect objects is a more stable and reliable cue than *-ko* with direct objects because *all* indirect objects are marked with *-ko* consistently whereas *some* direct objects are marked with zero. Therefore, input distribution also predicted higher accuracy on *-ko* marking of indirect object than on the other three markers.

The results of the AJT confirmed the trends observed in the oral task. The L2 learners and the heritage speakers accepted/produced more omission errors with ergative *-ne*, accusative *-ko* and dative experiencer *-ko* than with indirect objects *-ko*. There were also very few overgeneralization errors of *-ne* and *-ko*; in the case of *-ne* most overgeneralization errors were produced by the L2 learners [which suggests that they have difficulty with the application of impoverishment rule for ergativity in (10)]. Consistent with our hypothesis, the L2 learners and the heritage speakers did not omit case markers indiscriminately; variability was systematically constrained by the semantic complexity of features involved and distributional reliability of the case markers in the input.

Despite overall similar findings for the two groups, the heritage speakers exhibited more sensitivity to *-ko* use and omission than the L2 learners with inanimate direct objects, where *-ko* marking is preferred if the object is specific (DOM). Hindi does not have definite articles, and *-ko* marks definiteness and specificity. In the AJT, the heritage speakers were more accepting of *-ko* as a specificity marker with inanimate direct objects and of unmarked human non-specific objects than the L2 learners. The speakers from India and the heritage speakers produced and accepted *-ko* marking more than the L2 learners. We acknowledge that because the sentences were not presented in context, it was not possible to assess more directly whether these uses of *-ko* with inanimate non-specific objects were overgeneralization errors or renditions of specific objects, at least for the heritage speakers. These converging results from the two tasks suggest that heritage speakers seem to be more aware that *-ko* marks specificity than the L2 learners, which could be due to their longer exposure to the language since an earlier age. Further research should pursue the strength of this finding with tasks that manipulate context.

Except for the finding that heritage speakers seem to know that *-ko* marks specificity with direct objects better than the L2 learners, the reason why our heritage speakers were not more accurate on case in general may be related to the relatively advanced proficiency of the L2 learners and their exposure to reading and writing through instruction. Limited access to literacy affects heritage language development (Bayram et al., 2017). Laleko and Polinsky (2015), who investigated

knowledge of topic and case markers in Korean and Japanese, found that advanced L2 learners do as well as heritage speakers recognizing different case markers in these languages. Foote (2010) also found that high proficiency L2 learners do not differ from heritage speakers of Spanish in their production of agreement.

Summarizing so far, amount of exposure to English and Hindi and patterns of language dominance are two characteristics shared by the L2 learners and the heritage speakers. Although the three groups tested were bilingual in English and Hindi, including the Hindi speakers from India, the L2 learners and the heritage speakers grew up in an English-speaking environment, were residing in an English-speaking country and were English-dominant. Non-target mastery of case morphology in the two experimental groups favors the possibility that reduced exposure to Hindi in the United States results in fewer instances of case markers in the input. The pairing of form and meanings is not always transparent (i.e., opaque) in the Hindi case system because except for indirect objects, overt case marking with ergative *-ne*, accusative *-ko* and dative *-ko* with experiencers depends on syntactic and semantic conditions (see Table 2). As a result, acquiring case marking in Hindi requires sustained frequent input to learn the form-meaning mappings, which is what heritage speakers typically lack.

Finally, the dominant language itself may contribute to the morphological patterns found in the two English-dominant groups. Because English is not an ergative language and does not mark case overtly, it may also have contributed to difficulty marking case consistently in Hindi, the weaker language. On Keine's (2007) analysis of ergative *-ne* and accusative *-ko*, overt marking is the default and zero marking is the result of impoverishment rules, which predicts more errors of overgeneralization of *-ne* and *-ko* than omission. However, when the heritage speakers made errors, in general there were more omission than overgeneralization in the production task and no difference in the acceptability judgments between the two ungrammatical sentence types in the AJT. The L2 learners made both omission and overgeneralization errors, especially more overgeneralization errors with *-ne* in the oral production task, suggesting that they have yet fully acquired the impoverishment rule in (10). Perhaps the difference error patterns with ergatives may be related to the fact that L2 learners were receiving instruction and were more aware of ergativity marking than the heritage speakers, who were not receiving Hindi instruction. As for omission errors, or both L2 learners and heritage speakers, unmarked case in English is a very likely source of language influence and case omissions in Hindi. With the ergative, the heritage speakers and L2 learners may be reinforcing the nominative-accusative pattern from English in Hindi. A similar explanation for errors with ergative marking in minority language bilinguals has been advanced for Dyrbal in Australia (Schmidt, 1985) and Basque in Spain (Austin, 2007), where the ergative language is in contact with a nominative-accusative language. When *-ne* is present in Hindi, object agreement is with the object, not with the

subject. If the English-dominant bilinguals were assuming a nominative-accusative pattern for the two languages they would consistently produce agreement with the subject. We examined the production data and found that some heritage speakers omitted *-ne* and produced subject agreement but others produced object agreement. We also found ungrammatical cases of *-ne* omission and default agreement. In general, we did not observe any distinct pattern to support a change from ergative-absolutive to nominative-accusative. Our AJT did not include sentences testing ergativity marking with imperfective predicates and sentences manipulating different types of subject, object and default agreement errors. We acknowledge that this limitation prevents us from evaluating more directly transfer of the nominative-accusative system of English to the split ergative system of Hindi. This question requires a more in depth experimental study of ergativity and agreement patterns in Hindi as a heritage language and as a second language.

Finally, English dominance may also explain the omission of accusative *-ko* with animate specific direct objects as well as the omission of *-ko* with dative subjects, since English does not presently mark case overtly with direct objects or experiencer subjects. Spanish is a nominative-accusative language but like Hindi, it also marks animate specific indirect objects overtly. Studies on L2 Spanish (Guijarro Fuentes, 2012) and Spanish as a heritage language by speakers of English (Montrul, 1998, 2016b) have found high rates of case omission of the preposition “a” with animate objects and dative experiencer subjects. Therefore, the morphological variability found in the present study can easily be explained by dominant language transfer, in this case English, in the bilinguals with Hindi as weaker language.

## CONCLUSION

Our results confirmed that bilinguals with Hindi as non-dominant language show morphological variability and instability with morphological case, unlike fluent Hindi speakers who are balanced bilinguals. Case markers linked to semantic features like perfectivity of verbs, and animacy or specificity of nouns (ergative *-ne* and some instances of *-ko*) were more prone

to omission than case markers that are more predictable in the input, like the dative case of indirect objects. For both the L2 learners and the heritage speakers, the quantity and complexity of features bundled in morphemes coupled with amount of input and use of Hindi (reflected in overall proficiency) affect the strength of form-meaning mappings and their mastery of the case morphology. Morphological variability may be reinforced by knowledge of English, the dominant language in L2 learners and heritage speakers, which does not mark these cases overtly. AoA of Hindi may explain the heritage speakers' superior sensitivity to *-ko* as a specificity marker, but it did not modulate the overall level of case accuracy in this study. The acquisition and mastery of morphology in bilinguals seems to be determined more by amount of input and use than by age of onset of bilingualism and length of language use, especially in unbalanced bilinguals.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Review Board at the University of Illinois. The protocol was approved by the Office for the Protection of Human Subjects at the University of Illinois. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

SM designed the overall project, performed statistical analyses, and wrote the article. AB, RB, and VP created the Hindi version of the tests. AB and VP tested participants, transcribed, and coded oral data. AB and RB contributed to the writing as well.

## FUNDING

This study was partially funded by National Science Foundation (Grant No. BCS-0917593 to SM and RB).

## REFERENCES

- Abrahamsson, N., and Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Lang. Learn.* 59, 249–306. doi: 10.1111/j.1467-9922.2009.00507.x
- Aissen, J. (2003). Differential object marking: iconicity vs. economy. *Nat. Lang. Ling. Theory* 213, 435–448. doi: 10.1023/A:1024109008573
- Anand, P., and Nevins, A. (2006). “The locus of ergative case assignment: evidence from scope,” in *Ergativity: Emerging Issues*, eds A. Johns, D. Massam, and J. Ndayiragije (Berlin: Springer-Verlag), 3–26.
- Au, T. K., 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
- Austin, J. (2007). Grammatical interference and the acquisition of ergative case in bilingual children learning Basque and Spanish. *Bilingualism* 10, 315–331. doi: 10.1017/S1366728907003094
- Baten, K., Ponnet, A., and Verbeke, S. (2016). The acquisition of differential object marking in hindi as a foreign language. *Dutch J. Appl. Ling.* 5, 101–125. doi: 10.1075/dujal.5.2.01bat
- Baten, K., and Verbeke, S. (2015). “The acquisition of the ergative case in Hindi as a foreign language,” in *Theoretical and Methodological Developments in Processability Theory*, eds K. Baten, A. Buyl, K. Lochtmann, and M. Van Herreweghe (Amsterdam: John Benjamins), 71–104. doi: 10.1075/palart.4.04bat
- Bavin, E., and Stoll, S. (eds) (2013). *The Acquisition of Ergativity*. Amsterdam: John Benjamins. doi: 10.1075/tilar.9
- Bayram, F., Rothman, J., Iverson, M., Miller, D., Puig Mayenco, E., Kupisch, T., et al. (2017). Differences in use without deficiencies in competence: passives in the Turkish and German of Turkish heritage speakers in Germany. *Int. J. Biling. Educ. Biling.* 32, 1–21. doi: 10.1080/13670050.2017.1324403
- Bedore, L., Peña, E., Summers, C., Boerger, K., Resendiz, M., Greene, K., et al. (2012). The measure matters: language dominance profiles across measures



- in Spanish–English bilingual children. *Biling. Lang. Cogn.* 15, 616–629. doi: 10.1017/S1366728912000090
- Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Keynote article. Heritage languages and their speakers: opportunities and challenges for linguistics. *Theoret. Ling.* 39, 129–181.
- Bialystok, E. (1997). The structure of age. *Second Lang. Res.* 13, 116–137. doi: 10.1191/026765897677670241
- Bialystok, E., and Hakuta, K. (1999). “Confounded age. Linguistic and cognitive factors in age differences for second language acquisition,” in *Second Language Acquisition and the Critical Period Hypothesis*, ed. D. Birdsong (Mahwah, NJ: Lawrence Erlbaum), 161–181.
- Bittner, M., and Hale, K. (1996). Ergativity: toward a theory of a heterogeneous class. *Linguist. Inq.* 27, 531–604.
- Butt, M. (2006). *Theories of Case*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139164696
- Carreira, M., and Kagan, O. (2011). The results of the national heritage language survey: implications for teaching, curriculum design, and professional development. *Foreign Lang. Ann.* 44, 40–64. doi: 10.1111/j.1944-9720.2010.01118.x
- Chang, C. (2016). Bilingual perceptual benefits of experience with a heritage language. *Biling. Lang. Cogn.* 19, 791–809. doi: 10.1017/S1366728914000261
- Chang, C., Yao, Y., Haynes, E., and Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *J. Acoust. Soc. Am.* 129, 3964–3980. doi: 10.1121/1.3569736
- Chomsky, N. (2001). “Derivation by phase,” in *Ken Hale: A Life in Language*, ed. M. Kenstowicz (Cambridge, MA: MIT Press), 1–52.
- Chung, E.-S. (2016). Second and heritage language acquisition of Korean case drop. *Bilingualism* 21, 63–79. doi: 10.1017/S1366728916001218
- Coon, J. (2013). TAM split ergativity (Parts I–II). *Lang. Linguist. Compass* 7, 171–200. doi: 10.1111/lnc3.12011
- Davison, A. (2004). “Structural case, lexical case and the verbal projection,” in *Clause Structure in South Asian Languages*, eds V. Dayal and A. Mahajan (Amsterdam: Kluwer), 199–225. doi: 10.1007/978-1-4020-2719-2\_7
- Dayal, V. (2011). Hindi pseudo-incorporation. *Nat. Lang. Linguist. Theory* 29, 123–167. doi: 10.1007/s11049-011-9118-4
- de Swart, P., and de Hoop, H. (2007). “Semantic aspects of differential object marking,” in *Proceedings of SuB11*, ed. E. Puig-Waldmüller (Barcelona: Universitat Pompeu Fabra), 568–581.
- DeKeyser, R. (2013). Age effects in second language learning: stepping stones toward better understanding. *Lang. Learn.* 63(Suppl. 1), 52–67. doi: 10.1111/j.1467-9922.2012.00737.x
- DeKeyser, R., and Larson-Hall, J. (2005). “What does the critical period really mean?” in *The Handbook of Bilingualism. Psycholinguistic Approaches*, eds J. Kroll and A. de Groot (Oxford: Oxford University Press), 88–108.
- Dixon, R. (1994). *Ergativity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511611896
- Foot, R. (2010). Age of acquisition and proficiency as factors in language production: agreement in bilinguals. *Bilingualism* 13, 99–118. doi: 10.1017/S136672890999040X
- Goldschneider, J. M., and DeKeyser, R. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: a meta-analysis of multiple determinants. *Lang. Learn.* 51, 1–50. doi: 10.1111/1467-9922.00147
- Granena, G., and Long, M. (2013). Age of onset, length of residence, language aptitude, and ultimate attainment in three linguistic domains. *Second Lang. Res.* 29, 311–343. doi: 10.1177/0267658312461497
- Grosjean, F. (2008). *Studying Bilinguals*. Oxford: Oxford University Press.
- Guijarro Fuentes, P. (2012). The acquisition of interpretable features in L2 Spanish: personal a. *Biling. Lang. Cogn.* 15, 701–720. doi: 10.1017/S1366728912000144
- Halle, M., and Marantz, A. (1993). “Distributed morphology and the pieces of inflection,” in *The View From Building 20*, eds K. Hale and S. J. Keyser (Cambridge, MA: MIT Press), 53–109.
- Hansen, L. (1986). Universals in relative clause acquisition: evidence from child and adult L1 and L2 learners of Hindi-Urdu. *Lang. Learn.* 36, 143–158. doi: 10.1111/j.1467-1770.1986.tb00376.x
- Hawkins, R., and Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua* 118, 595–612. doi: 10.1016/j.lingua.2007.01.009
- 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
- Jia, G., and Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Appl. Psycholinguist.* 24, 131–161. doi: 10.1017/S0142716403000079
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Appl. Psycholinguist.* 25, 603–634. doi: 10.1017/S0142716404001298
- Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0
- Keine, S. (2007). “Reanalysing Hindi split-ergativity as a morphological phenomenon,” in *Linguistische Arbeits Berichte 85*, eds J. Trommer and A. Opitz (Leipzig: Universität Leipzig), 73–127.
- Kempe, V., and MacWhinney, B. (1998). The acquisition of case-marking by adult learners of Russian and German. *Stud. Second Lang. Acquis.* 20, 543–587. doi: 10.1017/S0272263198004045
- Kim, K., O’Grady, W., and Schwartz, B. (2016). Case in heritage Korean. *Linguistic Approaches to Bilingualism* 8, 252–282. doi: 10.1075/lab.16001.kim
- Lakshmanan, U. (1999). “Object shift and the position of NegP in the child L2 grammars of Hindi,” in *Representation and Process: Proceedings of the 1998 Pacific Second Language Research Forum*, Vol. 1, ed. P. Robinson (Tokyo: Pacific Second Language Research Forum), 23–36.
- Laleko, O., and Polinsky, M. (2015). Marking topic or marking case: a comparative investigation of Heritage Japanese and Heritage Korean. *Heritage Lang. J.* 10, 40–64.
- Laleko, O., and Polinsky, M. (2016). Between syntax and discourse: topic and case marking in heritage speakers and L2 learners of Japanese and Korean. *Linguist. Approach. Biling.* 6, 396–439. doi: 10.1075/lab.14018.lal
- Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Lang. Res.* 25, 173–227. doi: 10.1177/0267658308100283
- Lardiere, D. (2016). Missing the trees for the forest: morphology in second language acquisition. *Second Lang.* 15, 5–30.
- López, L. (2012). *Indefinite Objects*. Boston, MA: MIT Press. doi: 10.7551/mitpress/9165.001.0001
- Lukyanchenko, A., and Gor, K. (2011). “Perceptual correlates of phonological representations in heritage speakers and L2 learners,” in *Proceedings of the 35th Annual Boston University Conference on Language Development*, (Somerville, MA: Cascadia Press), 414–426.
- MacWhinney, B. (2008). “A unified model,” in *Handbook of Cognitive Linguistics and Second Language Acquisition*, eds P. Robinson and N. Ellis (Mahwah, NJ: Lawrence Erlbaum Associates).
- Marantz, A. (1991). “Case and licensing,” in *Proceedings of the ESCOL (East Conference on Linguistics)*, (Columbus, OH: Ohio State University), 234–253.
- Mohanan, T. (1993). Case Alternation on Objects in Hindi. *South Asian Lang. Rev.* 3, 1–30.
- Montrul, S. (1998). The L2 acquisition of dative experiencer subjects. *Second Lang. Res.* 14, 27–61. doi: 10.1191/026765898668810271
- Montrul, S. (2001). L1-constrained variability in the L2 acquisition of argument structure changing morphology with causative verbs. *Second Lang. Res.* 17, 144–194.
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Reexamining the Age Factor*. Amsterdam: John Benjamins. doi: 10.1075/sibil.39
- Montrul, S. (2011). Morphological errors in Spanish second language learners and heritage speakers. *Stud. Second Lang. Acquis.* 33, 155–161. doi: 10.1017/S0272263110000719
- Montrul, S. (2013). How “native” are heritage speakers? A look at gender agreement in Spanish. *The Heritage Language Journal* 10, 15–39.
- Montrul, S. (2016a). “Dominance and proficiency in second language acquisition and bilingualism,” in *Measuring Dominance in Bilingualism*, eds C. Silva-Corvalán and J. Treffers-Dallers (Cambridge: Cambridge University Press), 15–35.
- Montrul, S. (2016b). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139030502
- Montrul, S. (2018). “The bottleneck hypothesis and heritage language acquisition,” in *Meaning and Structure in Second Language Acquisition*, eds J. Cho, T. Judy,

- and T. Leal Mendez (Amsterdam: John Benjamins), 149–177. doi: 10.1075/sibil.55.06mon
- Montrul, S., Bhatt, R., and Bhatia, A. (2012). Erosion of case and agreement in Hindi heritage Speakers. *Linguist. Approach. Biling.* 2, 141–176. doi: 10.1075/lab.2.2.02mon
- Montrul, S., Bhatt, R., and Girju, R. (2015). Differential object marking in Spanish, Hindi, and Romanian as heritage languages. *Language* 91, 564–610. doi: 10.1353/lan.2015.0035
- Montrul, S. A., and Bowles, M. A. (2009). Back to basics: differential object marking under incomplete acquisition in Spanish heritage speakers. *Biling.* 12, 363–383. doi: 10.1017/S1366728909990071
- Montrul, S., Foote, R., and Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: the effects of age and context of acquisition. *Lang. Learn.* 58, 503–553. doi: 10.1111/j.1467-9922.2008.00449.x
- Montrul, S., and Gürel, A. (2015). “The acquisition of differential object marking in Spanish by Turkish speakers,” in *The Acquisition of Spanish by Speakers of Less Commonly Studied Languages*, eds T. Judy and S. Perpiñán (Amsterdam: John Benjamins), 281–308.
- Montrul, S., and Ionin, T. (2012). Dominant language transfer in Spanish heritage speakers and L2 learners in the interpretation of definite articles. *Mod. Lang. J.* 96, 70–94. doi: 10.1111/j.1540-4781.2012.01278.x
- Narasimhan, B. (2005). Splitting the notion of ‘agent’ case-marking in early child Hindi. *J. Child Lang.* 32, 787–803. doi: 10.1017/S0305000905007117
- Narasimhan, B. (2013). “Ergative case marking in Hindi child-caregiver speech,” in *The Acquisition of Ergativity*, eds E. Bavin and S. Stoll (Amsterdam: John Benjamins), 209–238.
- O’Grady, W. (2008). The emergentist program. *Lingua* 118, 447–464. doi: 10.1016/j.lingua.2006.12.001
- O’Grady, W., Kwak, H. Y., Lee, O.-S., and Lee, M. (2011). An emergentist perspective on heritage language acquisition. *Stud. Second Lang. Acquisit.* 33, 223–246. doi: 10.1017/S0272263110000744
- O’Grady, W., Lee, M., and Choo, M. (2001). The acquisition of relative clauses by heritage and non-heritage learners of Korean as a second language. A comparative study. *J. Korean Lang. Educ.* 12, 283–294.
- Pallotti, G. (2015). A simple view of linguistic complexity. *Second Lang. Res.* 31, 117–134. doi: 10.1177/0267658314536435
- Papadopoulou, D., Varlokosta, S., Spyropoulos, V., Kaili, H., Prokou, S., and Revithiadou, A. (2011). Case morphology and word order in second language Turkish: evidence from Greek learners. *Second Lang. Res.* 27, 173–205. doi: 10.1177/0267658310376348
- Paradis, J. (2010). 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
- Perpiñán, S. (2017). Catalan Spanish bilingualism continuum. The expression of non-personal Catalan clitics in the adult grammar of early bilinguals. *Linguist. Approach. Biling.* 7, 477–513. doi: 10.1075/lab.15004.per
- Polinsky, M. (2006). Incomplete acquisition: American Russian. *J. Slavic Linguist.* 14, 191–262.
- Putnam, M., and Sánchez, L. (2013). What’s so incomplete about incomplete acquisition? – A prolegomenon to modeling heritage language grammars. *Linguist. Approach. Biling.* 3, 478–508. doi: 10.1075/lab.3.4.04put
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: romance languages as heritage languages. *Int. J. Bilingual.* 13, 145–155. 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
- Santos, A., and Flores, C. (2016). Comparing heritage speakers and late L2 learners of European Portuguese. Verb-movement, VP-ellipsis and adverb placement. *Linguist. Approach. Biling.* 6, 308–340. doi: 10.1075/lab.14006.san
- Schmidt, A. (1985). *Young People’s Dyrbal. An Example of Language Death from Australia*. Cambridge: Cambridge University Press.
- Schwartz, B. D., and Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Lang. Res.* 12, 40–72. doi: 10.1177/026765839601200103
- Silva, R., and Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: evidence from masked priming experiments in English. *Biling. Lang. Cogn.* 11, 245–260. doi: 10.1017/S1366728908003404
- Silva-Corvalán, C. (1991). *Language Contact and Change*. New York: Oxford University Press.
- Silva-Corvalán, C. (2014). *Bilingual Language Acquisition: Spanish and English in the First Six Years*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139162531
- Silva-Corvalán, C., and Treffers-Dallers, J. (eds) (2016). *Measuring Dominance in Bilingualism*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107375345
- Slabakova, R. (2008). *Meaning in a Second Language*. Berlin: De Gruyter. doi: 10.1515/9783110211511
- 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
- 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
- Woolford, E. (2006). Lexical case, inherent case and argument structure. *Linguist. Inq.* 37, 111–130. doi: 10.1162/002438906775321175
- Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.
- Yeni-Komshian, G., Flege, J., and Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Biling. Lang. Cogn.* 3, 131–149. doi: 10.1017/S1366728900000225

**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 © 2019 Montrul, Bhatia, Bhatt and Puri. 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.



# Language Dominance and Cognitive Flexibility in French–English Bilingual Children

Elena Nicoladis\*, Dorothea Hui and Sandra A. Wiebe

Department of Psychology, University of Alberta, Edmonton, AB, Canada

## OPEN ACCESS

### Edited by:

Cornelia Hamann,  
University of Oldenburg, Germany

### Reviewed by:

Natalia Meir,  
University of Haifa, Israel  
Barbara Köpke,  
Université Toulouse - Jean Jaurès,  
France

### \*Correspondence:

Elena Nicoladis  
elenan@ualberta.ca

### Specialty section:

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

**Received:** 25 March 2018

**Accepted:** 22 August 2018

**Published:** 07 September 2018

### Citation:

Nicoladis E, Hui D and Wiebe SA  
(2018) Language Dominance  
and Cognitive Flexibility  
in French–English Bilingual Children.  
Front. Psychol. 9:1697.  
doi: 10.3389/fpsyg.2018.01697

Some studies have reported a cognitive advantage for bilingual children over monolinguals and other studies have not. One possible reason for these conflicting results is that the degree of cognitive flexibility is related to individual differences in language dominance and use. More balanced bilinguals who separate their languages by context might have to learn to reduce inter-language interference and therefore show greater cognitive flexibility. The goal of the present study was to test if language dominance is related to French–English bilingual children's cognitive flexibility, using three different measures of language dominance: (1) parental reports of dominance, (2) relative scores on vocabulary tests, and (3) knowledge of translation equivalents. We also included two measures of language use: (1) living in a bilingual community (Montreal) or a monolingual community (Edmonton) and (2) language separation. Sixty-two French–English bilingual between 46 and 85 months of age participated. Children's cognitive flexibility was assessed using the Advanced Dimensional Change Card Sort task. Children's language knowledge and use was assessed in both French and English using a battery of tests. The results showed that none of the measures of language dominance or language use predicted cognitive flexibility. These results are inconsistent with the claim that individual differences in language dominance and use predict bilinguals' executive functions.

**Keywords:** bilingualism, executive function, cognitive flexibility, translation equivalents, balanced bilingualism

## INTRODUCTION

Even when processing only one language, bilinguals have both languages activated in their minds (Grosjean, 1989; Green, 1998; Costa, 2005; Rodriguez-Fornells et al., 2005). Bilinguals therefore have to constantly control interfering information from their two active and competing language systems in order to select the relevant language and inhibit the other that is not in use at that moment (Rodriguez-Fornells et al., 2006; Abutalebi and Green, 2007; Costa et al., 2009; Festman and Münte, 2012). Some researchers have argued that bilinguals' experience with selecting and inhibiting languages could generalize to other tasks involving attentional processing and cognitive flexibility (Bialystok, 2001; Bialystok et al., 2005). Cognitive flexibility refers to the ability to shift between mental sets and tasks (Miyake et al., 2000) while selective attention refers to the ability to orient attention toward a specific stimulus while simultaneously ignoring other stimuli

(Plude et al., 1994; Mahone and Schneider, 2012). Both cognitive flexibility and selective attention are higher mental functions responsible for goal-directed behavior, or executive functions (EFs; Best and Miller, 2010).

If the experience of selecting and inhibiting languages generalizes to other tasks involving EF, bilinguals might outperform monolinguals on non-linguistic measures of EF. This bilingual advantage might be particularly salient during times of developmental change, like childhood and old age (Bialystok et al., 2006; Craik and Bialystok, 2006). During childhood, there is rapid growth of EFs during childhood due, in part, to the pronounced plasticity and maturation of the prefrontal cortex, which allows children to increasingly control their actions and thoughts (Diamond, 1991, 2002, 2009; Diamond and Taylor, 1996; Lengua et al., 2007; Conway and Stifter, 2012). This sensitive period of development provides an opportunity for some positive contextual experiences, such as socioeconomic status and parenting practices, to enhance the development of EFs (Matte-Gagne and Bernier, 2011; Sarsour et al., 2011; Fay-Stammach et al., 2014; Lengua et al., 2015).

Some research has shown that there are advantages of bilingualism on EF tasks involving conflict (Peal and Lambert, 1962; Bialystok, 1987, 2001, 2011; Bialystok and Martin, 2004; Bialystok et al., 2010; Poulin-Dubois et al., 2011; Barac and Bialystok, 2012; Garraffa et al., 2015; Antoniou et al., 2016; Blom et al., 2017; White, 2014, Unpublished). Conflict, in this context, is defined as a disagreement between two or more things, and it comes up whenever there are incompatible and competing responses or representations (Festman and Münte, 2012). Such tasks that present a conflicting situation include the dimensional change card sort task (DCCS). This task puts two pairs of rules in conflict with each other, and calls for children to pay attention to only one of them at a time (Bialystok, 1999).

Not all studies have shown a bilingual advantage on EF tasks (see review in Valian, 2015). Some recent studies found no differences in EFs between their monolingual and bilingual samples (Morton and Harper, 2007; Tare and Linck, 2011; Paap and Greenberg, 2013; Duñabeitia et al., 2014). In one study, the lack of difference held even when bilingual and monolingual children were carefully matched on age, gender, reading and mathematic skills, verbal and non-verbal IQ, family income, and the number of years of formal education of the parents (Antón et al., 2014). Other studies have found bilingual advantages only in particular age groups, such as preschoolers (Bialystok and Martin, 2004; Bialystok et al., 2006), children in middle childhood (Garraffa et al., 2015), young adults (Pelham and Abrams, 2014) or elderly adults (Craik and Bialystok, 2006; Bialystok et al., 2014). Other studies have not found a bilingual advantage in those age groups (see review in Valian, 2015).

One possible explanation for these variable findings is that there are individual differences among bilinguals that determine when an EF advantage is found (Festman et al., 2010; Valian, 2015). Indeed, some previous studies suggest that more balanced proficiency is associated with greater EF among bilinguals (Soveri et al., 2011; Yow and Li, 2015). The goal of this study was to investigate associations between bilingual children's language

dominance and use on the one hand and their EF abilities on the other.

## Language Dominance

One possible individual difference that could relate to bilingual children's degree of EF is their language dominance. Most bilingual children display more advanced knowledge or proficiency in one of their two languages, otherwise known as their dominant language (Paradis and Nicoladis, 2007; Gathercole, 2016). Language dominance is a complex and multi-faceted construct (Silva-Corvalán and Treffers-Daller, 2016). In the present study, we focus on dominance as it relates to proficiency in the two languages and use a variety of measures to assess the children's dominance. It is possible that more balanced bilinguals must reduce the interference from their other language more frequently than bilinguals with a strong dominance in one language. Indeed, some studies have shown that a more balanced knowledge in both languages leads to greater benefits on executive processes for bilinguals (Bialystok et al., 2006; Carlson and Meltzoff, 2008; Iluz-Cohen and Armon-Lotem, 2013; Tao et al., 2015; Antoniou et al., 2016; Blom et al., 2017; Thomas-Sunesson et al., 2018). For example, in a study comparing 6-year-old English-Italian bilinguals, English speakers with a limited knowledge of Italian, and English monolinguals, Ricciardelli (1992) found that the bilingual children performed significantly better than the other two groups in five of cognitive measures. Similarly, Bialystok and Majumder (1998) found that balanced bilingual children performed best on non-linguistic tasks that required control of attentional processing, even after controlling for age and language proficiency, when compared to the partial bilingual and monolingual groups. These patterns of results in the literature seem to suggest that the outcomes on cognitive performance may be dependent on the extent to which an individual is highly proficient in both languages.

Part of the reason that balanced bilingualism might be particularly strongly associated with cognitive advantages is the semantic organization of the mental lexicon. With increasing proficiency in both languages, bilingual children learn more translation equivalents (TEs) (Legacy et al., 2016). TEs refer to two words in different languages that refer to the same concept (e.g., *cake* and *gâteau*). Since both language systems are active in a bilingual's mind (Grosjean, 1989), there could be competition between TEs when a bilingual uses one of the words. A higher proportion of TEs would therefore mean that bilinguals have to switch across the two systems and inhibit the irrelevant one more frequently, thereby enhancing their EF (Patterson and Pearson, 2004). In support of this argument, a recent study by Crivello et al. (2016) found evidence of enhanced EF mechanisms as a function of TE acquisition, where a bigger increase in the number of TEs predicted higher EF in toddlers through increased opportunities for switching across the two lexical systems.

Some studies have not found a relationship between balanced proficiency and enhanced EFs (e.g., Paap et al., 2014; Tao et al., 2015; von Bastian et al., 2016). For example, in a large-scale study comparing English monolinguals with simultaneous and early sequential Welsh-English bilinguals on a variety of EF tasks, no bilingual advantage was found (Gathercole et al., 2014).



The authors attributed the lack of a bilingual advantage to the fact that their participants were all simultaneous or early sequential bilinguals. While this was a large-scale study and included participants from the age of 3 years through to adulthood, it is important to replicate. It is not clear why early onset of bilingualism would nullify a bilingual advantage, since early bilinguals, like late bilinguals, also show simultaneous activation of both languages even when processing only one (Nicoladis, 2006). Furthermore, some studies have shown EF advantages in both early and late bilinguals over monolinguals (Tao et al., 2011; Pelham and Abrams, 2014).

In sum, bilingual children's language dominance may predict their EF. Specifically, the more balanced bilinguals might have more experience inhibiting interference from the other language than less balanced bilinguals. Inhibiting interference from the other language may be particularly important for TEs, where there is competing activation from a word in the non-target language. Not all studies have found a relationship between dominance and EFs. It is possible that the link between balanced proficiency and enhanced EF does not hold for early-onset bilinguals.

## Language Use

Language selection decisions are thought to be important for EF gains among bilinguals (Rubin and Meiran, 2005; Festman et al., 2010; Soveri et al., 2011). However, some studies of adult bilinguals have shown no relationship between language usage and EFs (Paap et al., 2014; von Bastian et al., 2016). In the studies with adults, a common measure is the frequency of usage of the two languages on a day-to-day basis.

In the present study, we measured how separately the bilingual children kept their two languages with monolingual interviewers in their two languages. While bilingual children can be shown to differentiate their two languages from early in development (Arnberg and Arnberg, 1992; Mishina-Mori, 2002; see review in Quay and Montanari, 2016), there is variability in how separately the languages of bilinguals are kept in use. While language separation might be at least somewhat related to language dominance, the two constructs can at least sometimes diverge. For example, a recent study showed that Chinese–English bilingual children could use as many different word types in their second language English as English monolinguals even though their proficiency in English was much weaker than that of the same-aged monolinguals (Nicoladis and Jiang, 2018). Furthermore, some bilinguals live in bilingual communities in which people use both languages with multiple people on an everyday basis, while other bilinguals live in more monolingual regions, requiring a greater degree of separation in use (Poplack, 1985; Ayeomoni, 2006; Baker, 2011; Gatt et al., 2016). In this study, we test whether children use a high degree of the target language in interviews in both of their languages.

One measure of the degree of separation in use is code-switching, or the use of two languages in a single unit of discourse (MacSwan, 2016). For example, a Spanish–English bilingual might say, *me voy al mall* 'I'm going to the mall', with most words in Spanish and the English word *mall*. In some studies, the frequency of code-switching among bilingual

children has been found to be related to their proficiency (Nicoladis and Genesee, 1996; Ribot and Hoff, 2014; Yow et al., 2017), but not all studies have found the same directionality of that relationship. Ribot and Hoff (2014) showed that when the target language was Spanish, Spanish–English bilinguals in the United States used more English as their English proficiency increased. Thus, the children kept the languages less separate as their English proficiency increased. In contrast, for French–English bilingual children in Montreal, their use of code-switching largely reflected their proficiency in the target language, with higher proficiency in the target language being associated with lower use of code-mixing (Nicoladis and Genesee, 1997). In other words, the French–English bilingual children could have been code-switching to fill gaps in their knowledge in the target language. These results suggest that, in some communities, like Montreal, separation in language use may simply reflect a high degree of proficiency in both languages. In others, a high degree of proficiency in both languages could result in a low use of code-switching, particularly to the majority language of the community.

The community language practices could also be related to EFs in bilinguals. Tao et al. (2015) found a link between dominance and EF in Spanish–English bilinguals but not Mandarin–English bilinguals. Their study took place in Southern California where Spanish is a commonly used language. One explanation they considered for the different results among the two bilingual groups was that the groups differ on how frequently they have to monitor switches between two languages. Since there are many Spanish speakers in Southern California and many monolingual English speakers, Spanish–English bilinguals might have to do a lot of monitoring for which language is appropriate in a particular instance. In contrast, since there are few Mandarin speakers in Southern California, Mandarin–English bilinguals might have fewer instances of selecting an appropriate language. To our knowledge, the possible effect of the language community has not been taken into account in studies of EFs in bilingual children.

The present study was conducted with Canadian French–English bilinguals in both Edmonton, Alberta, and in Montreal, Quebec. In Montreal, many of its residents are fluent in both French and English (see Sioufi et al., 2016, for arguments classifying Montreal as part of Canada's "bilingual belt"). Edmonton is a majority English-speaking city with a small francophone minority population (Aunger, 1999). Children's language use shows effects of the linguistic community in the preschool years. For example, Genesee et al. (1995) showed that in Montreal French–English bilingual children's code-switching, the use of two languages in a single unit of discourse, was related primarily to their dominant language (see also Nicoladis and Genesee, 1997). Specifically, the greater their proficiency in the language, the less likely they were to use words from their weaker language. In contrast, Paradis and Nicoladis (2007) showed that the code-switching among French–English bilingual children in Edmonton was affected not only by their stronger language but also by which language was the majority one. That is, both English-dominant and French-dominant children code-mixed infrequently in English (see also Ribot and Hoff, 2014). Paradis and Nicoladis (2007) argued that the Edmontonian children

were sensitive to the fact that most French speakers also speak English while not all English speakers reliably speak French.

Bilinguals in Edmonton may therefore have to develop greater EF to avoid code-switching and interference from the non-relevant language on a daily basis. This enhanced practice in language control and language separation may in turn translate to larger EF advantages for the Edmonton bilinguals relative to the Montreal bilinguals.

In sum, bilingual children vary in how separate they keep their two languages in use. The degree of separation in use may be related to the children's proficiency in their two languages as well as the community in which children live. If children keep their two languages separate in use, they may have greater EF than if they do not. Children living in an English-majority-language community like Edmonton might have greater practice keeping the two languages separate than children living in a bilingual community like Montreal. The children from a monolingual community might therefore show higher EF than the children from a bilingual community.

## Research Questions

The purpose of this study was to test language dominance and language use predictors of bilingual children's cognitive flexibility. We included three measures of language dominance: (1) parental report, (2) relative vocabulary scores, and (3) TEs. The rationale for including three measures of dominance is that previous research has shown that different measures of dominance can yield different results (Bedore et al., 2012). We predicted that the more balanced the bilinguals and the higher rate of TEs, the higher their EF would be. We included two measures of language use: (1) linguistic community and (2) degree of language separation. We predicted that higher EF would be observed among the children in Edmonton (English majority-language community) relative to the children in Montreal (bilingual community) and among children who kept their languages separate in use relative to those who did not. These predictions are based on the assumption that the greater degree of experience inhibiting an inappropriate language, the greater executive control they would have to exercise on a daily basis, and therefore the more successful we expected them to be in the EF task.

The design of this study was correlational. Therefore, if we find the predicted correlations, we cannot identify directionality. We have phrased our research questions as if it is the experience with learning and separating languages that leads to enhanced EFs. However, it is equally possible that it is children with enhanced EFs who are more likely to become balanced bilinguals and separate their languages in use (Festman et al., 2010). We return to this point in the discussion.

## MATERIALS AND METHODS

### Participants

The sample included a total of 62 French–English bilingual children, 36 from Montreal and 26 from Edmonton. The group comprised of 29 boys and 33 girls and had an age range of

46–82 months ( $M = 59.44$ ,  $SD = 8.04$ ). With regards to the age of exposure to both languages: 48 children were reported to have been exposed to both languages since birth, 10 between 1–2 years of age, and four between 2–4 years of age. The four bilinguals with age of onset to one language between 2–4 years were not outliers within the groups on any of our measures and so were included in all analyses.

The parents were asked: “Please choose the best description of your child's French/English knowledge.” They were given five choices: (a) My child speaks French far better than English, (b) My child speaks French a bit better than English, (c) My child speaks both languages about equally well, (d) My child speaks English a bit better than French, and (e) My child speaks English far better than French. According to parental report, 18 were relatively balanced (i.e., chose option c; 11 were girls), 18 were slightly dominant (9 in French [option b] and 9 in English [option d]; 7 girls) and 25 were strongly dominant (11 in French [option a] and 14 in English [option e]; 14 girls). One parent did not respond to this question; this child's data were excluded from the analyses including this measure.

## Procedure

The current study obtained approval from the institutional research ethics board. Parents signed an informed consent form, giving us permission to test their children. All children were asked for verbal assent before any tasks were carried out. The children completed a battery of language and cognitive tasks on different days for the two languages. We present the results only of the measures that are related to our research questions here. The order of the tests with a testing session varied according to the child's engagement and comfort with the experimenter. As a default, the more passive tasks (such as the receptive vocabulary test) were administered earlier in the sessions than the more active tasks (such as the story-telling task). The order of the French and English sessions was counter-balanced. Different experimenters ran the sessions and both were native speakers of the testing language.

## Measures

### Vocabulary

The Peabody Picture Vocabulary Test III (PPVT; Dunn and Dunn, 1997) and the Echelle de Vocabulaire en Images Peabody (EVIP; Dunn et al., 1993 – the French version of the PPVT) were used to measure children's English and French receptive vocabulary size. Children had to respond to single words spoken aloud by the experimenter by either pointing to or indicating the number of the appropriate picture out of four black-and-white pictures. In accordance with PPVT standard starting and stopping criteria, the task started at children's PPVT age set and stopped when children identified 8 or more items in a given set incorrectly.

The raw scores for the PPVT and EVIP are not on the same scale. For example, one 62-month old boy received a raw score of 91 on the PPVT and a raw score of 72 on the EVIP while his standard scores were virtually identical in the two languages (i.e., 118 on the PPVT and 120 on the EVIP). Therefore, in order to determine the child's vocabulary dominance, we used the

standard scores. For each child, we first noted in which language the standard score was higher. We then calculated the ratio of the standard scores (with the higher standard score for the child divided by the lower standard score for the child). Thus, the closer to one, the more balanced vocabulary scores a child showed. A higher ratio means that the children were showing greater unbalance in their vocabulary scores.

### Language Separators

Children were shown a Pink Panther video, and then asked to recount as many details about it as they could after watching it. The children did this in both French and English. The total number of words in the stories told were calculated, including the number of French words used during the English session and the number of English words used during the French session (see **Table 1** for summary statistics). All the children but one used 92% or more words in one language or another in both sessions (the one exception used 81% French words in the French session). Given the categorical nature of children's behavior, we classified the children as either language separators or not based on how much of the target languages the children used for both languages. To be classified as a language separator, a child had to use 92% or more French words in the French session and 92% or more English words in the English session. A child was classified as not being a language separator if his/her language use was less than 92% of the target language in one language. No child used less than 92% of the target language in both languages. Six children (4 balanced, 1 slightly unbalanced, and 1 unbalanced according to the parental report) were not included in this classification because they were not videotaped in one language or the other due to scheduling conflicts.

### Translation Equivalents

The verbal semantic fluency task was used as a measure for children's TEs. In this test, children were asked to name words from the following categories: clothes, animals, and food plus drinks. The given time per category was 30 s. The children did this in French during the French session and in English during the English session. The score obtained was a percentage of words that were TEs out of the total number of concepts generated. For example, if the child said "cat, dog" in English and "chat, grenouille" 'cat, frog' in French, the

child generated a total of three concepts (cat, dog, and frog) with only one TE (for the concept cat). The ratio of TEs would be  $1/3 = 0.333$ . We then multiplied by 100 to make a percentage.

Verbal semantic fluency tasks can measure both lexical knowledge and lexical retrieval (Weckerly et al., 2001) as well as executive control ability (Ruff et al., 1997; Shao et al., 2014). The total number of correct words generated is related to language ability, especially word knowledge, like vocabulary size (Ruff et al., 1997; Sergeant et al., 2002). The order in which words are generated can reflect executive functioning (Troyer et al., 1997; Hurks et al., 2010). In the present study we focused exclusively on the words generated rather than the order in which they were generated. This measure should therefore show strong correlations with the other dominance measures.

### Executive Function Task

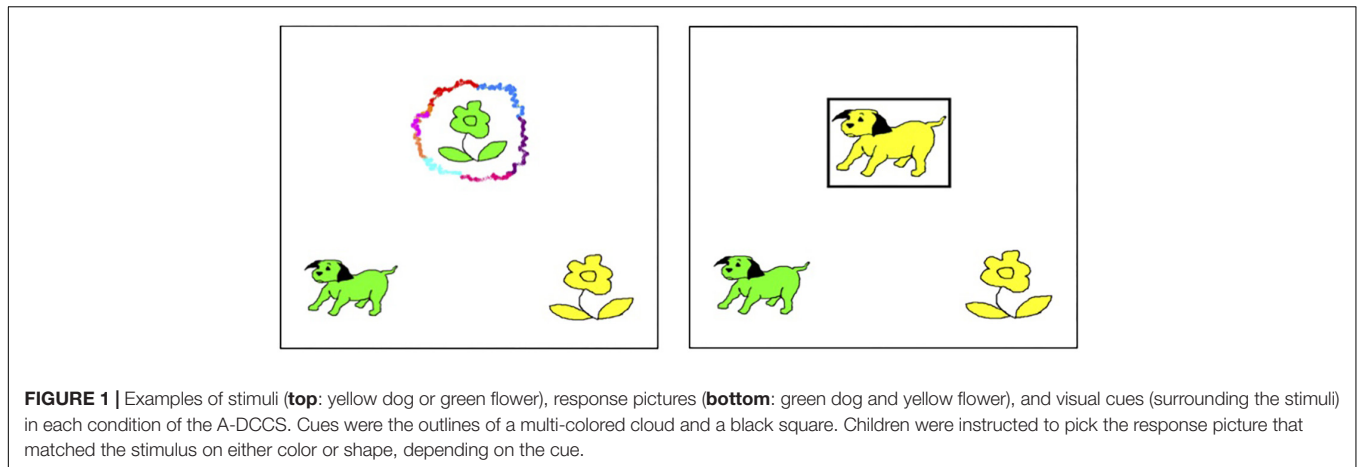
The Advanced Dimensional Change Card Sorting (A-DCCS) task (adapted from Chevalier and Blaye, 2009) was used as an index of participants' EF. The task was run with E-Prime 2.0 software (Psychology Software Tools, Inc., 2007) and administered on a laptop computer during the English session. Children were asked to respond as quickly and accurately as possible by pressing either the "q" or "p" keys on the laptop keyboard, with the remaining keys covered and masked. In this computerized task, participants were required to match, based on task cues, a stimulus with one of two response pictures on either shape (Shape Game) or color (Color Game) on each trial. Stimuli were two pictures of different shapes and colors (a green flower and a yellow dog) and were presented at the top of the screen in the center (**Figure 1**). Each response image, a yellow flower and a green dog, matched the two stimuli on either shape or color. The two response images remained on the screen throughout the task and were presented on the two bottom corners of the screen and corresponding with the "p" and "q" keys respectively (**Figure 1**). Task cues surrounding the stimuli to indicate which game children should play were a multi-colored cloud (Color Game) and a black square (Shape Game; see **Figure 1**).

The A-DCCS consisted of three phases: color, shape, and mixed. Each phase started with a practice block followed by

**TABLE 1** | Mean (SD) scores on the predictor variables by parental reports of dominance group.

	Balanced ( <i>N</i> = 18)	Slightly dominant ( <i>N</i> = 18)	Very dominant ( <i>N</i> = 25)
Age in months	60.9 (7.0)	65.6 (9.3)	56.4 (4.8)
PPVT, standard scores	101.0 (14.5)	99.8 (15.1)	90.1 (30.6)
EVIP, standard scores	107.3 (17.1)	102.4 (17.4)	94.6 (24.4)
Vocabulary ratio	9.1 (7.3)	17.6 (17.0)	38.5 (32.3)
%Translation equivalents	23.5% (10.9%)	15.6% (12.6%)	8.7% (8.1%)
#Montreal/Edmonton	10/8	13/5	10/15
Language separator (# Yes/No) <sup>†</sup>	14/1	14/3	7/17
Average % Code-mixed words in English story	0% (0%)	0% (0%)	10.4% (31.1%)
Average % Code-mixed words in French story	0.2% (0.6%)	7.8% (23.1%)	44.8% (50.0%)

PPVT, English receptive vocabulary; EVIP, French receptive vocabulary. <sup>†</sup>Numbers do not necessarily total because of missing data (see text).



one test block for the color and shape phases, and two test blocks for the mixed phase. The experimenter provided children with help during practice blocks if necessary, but not during the test blocks. On every trial, the task cue and stimulus were simultaneously presented. Once a response was entered, the stimulus (without the task cue) would move onto the side of the given response. This was to simulate the traditional card version of the A-DCCS where cards are put into boxes, making the button-press response real for children. In the Color Game, children were instructed to press the key under the bottom (response) picture of the same color as the top (stimulus) picture. For the Shape Game, children were instructed to press the key under the response picture of the same shape as the stimulus. In the mixed phase, children were told that they would be playing both games at the same time. The mixed phase contained non-switch trials – where the relevant game recurs – and switch trials – where the relevant game changes.

The dependent variables for this task were the mixing and switching costs in accuracy, as well as reaction time for switching costs. However, it is important to note that previous research found a bilingual advantage in only the switching cost (Garbin et al., 2010; Prior and MacWhinney, 2010). We nonetheless include the mixing costs to verify the generalizability of the previous null findings. For all the reaction time measures, we removed the data for any child who responded two or more SDs slower than the average for that particular measure (between 3 and 5% of the data). Mixing costs compare performance on simple phases (just one relevant task/game across the entire block) with non-switch trials (the relevant task/game is identical to that of the previous trial) from the mixed phases (both tasks/games are relevant within the block; Chevalier and Blaye, 2009). Switching costs, on the other hand, compare switch trials (the relevant task/game is different from the previous trial) and non-switch trials within the mixed phases only. These are used in the literature as comparable to the bilingual process of juggling two languages: mixing costs reflect the task-decision process of goal setting and the difficulty of keeping two task sets activated; while switching costs reflect the switching process (Rubin and Meiran, 2005; Chevalier and Blaye, 2009).

Goal setting is primarily reflected in mixing costs as it is present in the non-switch trials but not in the simple blocks (Chevalier and Blaye, 2009). As for the switching costs, both the non-switch and switch trials need goal setting, but only the latter require implementing a switch (Chevalier and Blaye, 2009). The dependent variables with accuracy were calculated using the following equations:

Mixing costs in accuracy =

$$\text{Single accuracy} - \text{Mixed non} - \text{switch accuracy} \quad (1)$$

Switching costs in accuracy =

$$\text{Mixed non} - \text{switch accuracy} - \text{Mixed switch accuracy} \quad (2)$$

We also included the reaction time for the switching costs, since Blom et al. (2017) found that some children showed a negative Flanker effect in reaction times.

## RESULTS

Descriptive statistics for the predictor variables and the EF measures are summarized in **Table 2**, grouped by the parental report on the children's dominance. The children in each dominance group were not equivalent on age (see **Table 1**),  $F(2,59) = 9.39$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.241$ . LSD *post hoc* comparisons showed that all of the dominance groups differed from each other at  $p \leq 0.049$ . The slightly dominant group was the oldest on average, followed by the balanced group, followed by the very dominant group. Given the age differences between groups, we partialled out age in presenting the main analyses in **Table 3**. As can be seen in **Table 2**, some children showed similar negative cost effect as reported in Blom et al. (2017), with many children responding faster to mixed switch trials than to mixed non-switch trials ( $N = 24$ ).

We had predicted that the parental reports of dominance, the vocabulary ratio, and the percentage of TEs (%TEs)



**TABLE 2 |** Scores on the advanced dimensional change card sort task by parental reports of dominance group.

	Balanced ( <i>N</i> = 18)	Slightly dominant ( <i>N</i> = 18)	Very dominant ( <i>N</i> = 25)
Single accuracy	0.92 (0.12)	0.87 (0.12)	0.89 (0.09)
Mixed non-switch accuracy	0.81 (0.16)	0.77 (0.17)	0.77 (0.17)
Mixed switch accuracy	0.79 (0.13)	0.71 (0.17)	0.72 (0.15)
Single RT	1545.4 (596.4)	1470.6 (713.1)	1961.2 (998.1)
Mixed non-switch RT	2638.0 (1207.4)	2583.6 (707.9)	3782.0 (1255.4)
Mixed switch RT	2804.2 (1340.0)	3181.5 (1635.0)	4488.8 (1909.1)
Mixing costs in accuracy	0.09 (0.19)	0.10 (0.21)	0.19 (0.19)
Switching costs in accuracy	0.08 (0.12)	0.03 (0.12)	0.05 (0.14)
Switching costs in reaction time	26.9 (638.4)	382.5 (1900)	−4.1 (4824.3)

**TABLE 3 |** Correlations between age, dominance measures, language use measures, mixing costs in accuracy, switching costs in accuracy, and switching costs in reaction times.

Variables	Parental dominance	Vocabulary ratio	TEs	Separators	City	Mixing costs	Switching costs	Reaction times
Age	−0.27*	−0.07	0.24	0.30*	0.02	−0.14	−0.06	−0.07
Parental dominance	—	0.50**	−0.52**	−0.57**	0.15	0.24	−0.07	−0.01
Vocabulary ratio	0.44**	—	−0.32*	−0.27	0.18	0.00	−0.03	0.11
%TEs	−0.46**	−0.34*	—	0.50**	−0.14	−0.01	−0.14	0.06
Language separators	−0.55**	−0.25	0.43**	—	−0.17	−0.22	−0.05	−0.12
City	0.03	0.24	−0.09	−0.18	—	0.11	0.13	0.15
Mixing costs in accuracy	0.17	−0.02	0.01	−0.25	0.07	—	−0.25*	−0.25
Switching costs in accuracy	−0.06	−0.03	−0.08	0.02	0.15	−0.07	—	0.07
Switching costs in reaction times	−0.03	0.10	0.07	−0.14	0.15	−0.26*	0.07	—

Shaded cells (below the diagonal) show correlations with age partialled out. Parental dominance = parental report of dominance; %TEs = percent translation equivalents; language separators dummy coded so that 0 = no and 1 = yes; city dummy coded so that 0 = Montreal and 1 = Edmonton. Partial correlations for age below the diagonal. \**P* < 0.05, \*\**P* < 0.01.

would be converging measures of language dominance. The first-order correlations between the variables under study are summarized above the diagonal in **Table 3**. To include City as a correlate, Edmonton was coded as 0 and Montreal as 1. As can be seen in this Table, all the dominance measures are all highly intercorrelated in the predicted direction. For example, the negative correlation between parental dominance and TEs means that the more balanced the parents judged their child to be, the more TEs that child produced. One measure of language use (language separators) was also highly correlated with the dominance measures, such that the more balanced the children, the more likely they were to separate their languages in use. The other measure of hypothesized language use (City) was not related to any of the other variables. The language dominance and use variables did not correlate significantly with the mixing costs in accuracy, the switching costs in accuracy, or reaction times on the A-DCCS.

Below the diagonal in the shaded cells of **Table 3**, we present the correlations between variables, with age partialled out. It is the last three rows of this Table that present the data to address our research questions. None of the language dominance or language use measures was significantly correlated with mixing cost accuracy, switching cost accuracy, or reaction times for switching costs. In contrast, many of the language dominance and use measures remained

significantly correlated with each other, even after controlling for age.

## DISCUSSION

The purpose of the present study was to test whether language dominance and language use measures predicted the EF of French–English bilingual children. We used multiple measures of language dominance: parental report, the ratio of standardized vocabulary scores, and the percentage of TEs generated on a semantic verbal fluency task. We measured language use both by linguistic community (the English-majority-language community of Edmonton vs. the bilingual community of Montreal) and by whether the children separated their languages by the language of the interlocutor. We predicted that the children who were relatively balanced in their bilingual abilities would perform better on the EF task than the unbalanced bilinguals (Ricciardelli, 1992; Bialystok and Majumder, 1998). The task-decision process that happens in the mixed-task blocks resembles the bilingual situation where decisions of which language to use have to be made in conversations, and smaller mixing costs reflect better control of attentional processing and higher ability in keeping two different task-sets active (Braver et al., 2003; Soveri et al., 2011; Brocki and Tillman, 2014). Consequently, there is reason to expect that balanced bilinguals

have more experience in having both languages activated and paying attention to non-salient features of input (Bialystok and Majumder, 1998).

Our results showed no relationship between either language dominance or language use measures and children's EF as indexed by mixing or switching costs in accuracy or reaction times for switching costs. These results contrast with those of some previous studies, showing larger EF advantages for balanced bilinguals (e.g., Ricciardelli, 1992; Bialystok and Majumder, 1998; Crivello et al., 2016). In some previous studies, a bilingual advantage has been shown in only switching costs (Garbin et al., 2010; Prior and MacWhinney, 2010). In this study, we found no correlation between either switching or mixing costs and language dominance or use.

One possible reason for the lack of relationship between the language dominance and use measures with EF is that the participants in the present study were either simultaneous or early sequential bilinguals. Recall that Gathercole et al. (2014) argued that they did not show a bilingual advantage because their participants were simultaneous or early sequential bilinguals. We think this is an unlikely explanation for two reasons. First, other studies with early-onset bilinguals have shown advantages (Bialystok and Majumder, 1998). Second, early bilinguals also show simultaneous activation of both of their languages (Grosjean, 2010).

Another possible reason for the null findings is that we did not have enough statistical power to show a significant relationship between the language measures and EFs. A power analysis showed that we have 80% power to detect a correlation of 0.35 (two-sided), suggesting that we do have adequate power. Also, other studies showing positive effects have sometimes included smaller or equivalent sample sizes than the ones here (e.g., Crivello et al., 2016; Thomas-Sunesson et al., 2018). Furthermore, studies including even larger sample sizes have shown null effects (e.g., Gathercole et al., 2014).

A third, and we expect the most likely, possible explanation for our results is the following. If there are individual differences between bilingual children that predict the degree of EF, language dominance and use may not be valid predictors. Other researchers have raised other possible variables that could contribute including socioeconomic status (Morton and Harper, 2007; cf. Kang et al., 2016), immigration status (de Bruin et al., 2015), culture (Kang et al., 2016; cf. Barac and Bialystok, 2012), working memory capacity (Namazi and Thordardottir, 2010), and others (see Donnelly et al., 2015).

If this explanation is correct, then there is growing evidence that the rationale behind predicting a bilingual advantage in EF needs to be reconsidered. Note that the present study was not designed to test whether there is a bilingual advantage, an issue that has been addressed extensively elsewhere (e.g., Paap and Greenberg, 2013; Paap et al., 2015; Sorge et al., 2017). Rather, the purpose of this study was to test whether the degree of language dominance or separation in use predicted EF. The rationale for a bilingual advantage has been that the experience selecting the appropriate language for the context and inhibiting the inappropriate

language would lead to general EF advantages. We found no evidence for this claim (consistent with some other studies; Gathercole et al., 2014; Paap et al., 2014; von Bastian et al., 2016).

We noted at the outset of our study that our correlational design does not allow us to distinguish the directionality of effects (or lack of effects). Some researchers have argued that bilinguals with high EF ability may be the ones who become highly proficient in both languages and/or learn to separate the languages well (e.g., Festman and Münte, 2012). While we see no evidence for this interpretation in our present study, we should also point out that our study was not designed to test that prediction. Some studies have found a bilingual advantage in older children but not younger children (e.g., Garraffa et al., 2017; although cf. Gathercole et al., 2014). A better design to test the possibility that high EFs lead to balanced proficiency and use would be a longitudinal one.

Before closing, we would like to draw readers' attention to one unexpected finding. The linguistic community (either monolingual Edmonton or bilingual Montreal) was not correlated with any of the other measures of language dominance and use. In contrast, the other measures of language dominance and use tended to be highly intercorrelated. Recall that we had predicted that Edmonton bilinguals would be more likely to keep their two languages separate in use than Montreal bilinguals. One possible reason for this finding is that an individual child's linguistic community may vary from the larger community (see Byers-Heinlein et al., 2017, for discussion specifically about children's individual language communities in Montreal). Previous studies have shown that family language practices can affect bilingual children's language dominance and use (e.g., Altman et al., 2014; see review in Quay and Montanari, 2016). Future research can test for that possibility.

## CONCLUSION

The present study showed no relationship between bilingual children's cognitive flexibility and language dominance/use. These results, in combination with others, raise doubts as to the rationale usually given for a purported bilingual advantage in EF. To the extent that there is a bilingual advantage in non-linguistic EF tasks, it may not be because of experience selecting and inhibiting languages alone, at least within the age range we examined here. Other researchers have raised other possibilities, including language proximity (e.g., Antoniou et al., 2016; Garraffa et al., 2017; cf. Antón et al., 2014) or task specificity (see review in Valian, 2015) or developmental changes (Garraffa et al., 2017).

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Tri-Council policy of Canada. The protocol was approved by the Research Ethics Board of the

University of Alberta. The parents or legal guardians of all participants gave written informed consent.

## AUTHOR CONTRIBUTIONS

EN supervised the data collection, and helped with data analysis and writing the manuscript. DH ran the first analyses, wrote the first draft of the manuscript, and provided feedback on the

penultimate draft. SW provided the appropriate EF measure, and contributed to the analyses and write-up of the manuscript.

## FUNDING

EN received funding for this study from the Natural Sciences and Engineering Research Council of Canada (Discovery Grant # 239851).

## REFERENCES

- Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguist.* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003
- Altman, C., Burstein Feldman, Z., Yitzhaki, D., Armon Lotem, S., and Walters, J. (2014). Family language policies, reported language use and proficiency in Russian–Hebrew bilingual children in Israel. *J. Multiling. Multicult. Dev.* 35, 216–234. doi: 10.1080/01434632.2013.852561
- 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
- Antoniou, K., Grohmann, K. K., Kambanaros, M., and Katsos, N. (2016). The effect of childhood bilingualism and multilingualism on executive control. *Cognition* 149, 18–30. doi: 10.1016/j.cognition.2015.12.002
- Arnberg, L. N., and Arnberg, P. W. (1992). “Language awareness and language separation in the young bilingual child,” in *Cognitive Processing in Bilinguals*, ed. R. J. Harris (Amsterdam, NH: Elsevier Science Publishers), 475–500.
- Aunger, E. (1999). “Les communautés francophones de l’Ouest: la survivance d’une minorité dispersée dans,” in *Francophonies minoritaires au Canada : L’état Des Lieux* ed. J.Y. Thériault (Moncton, Éditions d’Acadie), 283–304.
- Ayeomoni, M. O. (2006). Code-switching and code-mixing: style of language use in childhood in Yoruba speech community. *Nordic J. Afr. Stud.* 15, 90–99.
- Baker, C. (2011). *Foundations of Bilingual Education and Bilingualism*. Toronto, ON: Multilingual Matters.
- Barac, R., and Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: role of language, cultural background, and education. *Child Dev.* 83, 413–422. doi: 10.1111/j.1467-8624.2011.01707.x
- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., et al. (2012). The measure matters: language dominance profiles across measures in spanish–english bilingual children. *Bilingualism* 15, 616–629. doi: 10.1017/S1366728912000090
- Best, J. R., and Miller, P. H. (2010). A developmental perspective on executive function. *Child Dev.* 81, 1641–1660. doi: 10.1111/j.1467-8624.2010.01499.x
- Bialystok, E. (1987). Influences of bilingualism on metalinguistic development. *Second Lang. Res.* 3, 154–166. doi: 10.1177/026765838700300205
- 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. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511605963
- Bialystok, E. (2011). Coordination of executive functions in monolingual and bilingual children. *J. Exp. Child Psychol.* 110, 461–468. doi: 10.1016/j.jecp.2011.05.005
- Bialystok, E., Barac, R., Blaye, A., and Poulin-Dubois, D. (2010). Word mapping and executive functioning in young monolingual and bilingual children. *J. Cogn. Dev.* 11, 485–508. doi: 10.1080/15248372.2010.516420
- Bialystok, E., Craik, F. I. M., and Ruocco, A. C. (2006). Dual-modality monitoring in a classification task: the effects of bilingualism and ageing. *Q. J. Exp. Psychol.* 59, 1968–1983. doi: 10.1080/17470210500482955
- Bialystok, E., and Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Appl. Psycholinguist.* 19, 69–85. doi: 10.1017/S0142716400010584
- Bialystok, E., and Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Dev. Sci.* 7, 325–339. doi: 10.1111/j.1467-7687.2004.00351.x
- Bialystok, E., Martin, M. M., and Viswanathan, M. (2005). Bilingualism across the lifespan: the rise and fall of inhibitory control. *Int. J. Biling.* 9, 103–119. doi: 10.1111/j.1467-7687.2008.00746.x
- Bialystok, E., Poarch, G., Luo, L., and Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychol. Aging* 29, 696–705. doi: 10.1037/a0037254
- Blom, E., Boerma, T., Bosma, E., Cornips, L., and Everaert, E. (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Front. Psychol.* 8:552. doi: 10.3389/fpsyg.2017.00552
- Braver, T. S., Reynolds, J. R., and Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron* 39, 713–726. doi: 10.1016/S0896-6273(03)00466-5
- Brocki, K. C., and Tillman, C. (2014). Mental set shifting in childhood: the role of working memory and inhibitory control. *Infant Child Dev.* 23, 588–604. doi: 10.1002/icd.1871
- Byers-Heinlein, K., Behrend, D. A., Said, L. M., Giris, H., and Poulin-Dubois, D. (2017). Monolingual and bilingual children’s social preferences for monolingual and bilingual speakers. *Dev. Sci.* 20:e12392. doi: 10.1111/desc.12392
- 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
- Chevalier, N., and Blaye, A. (2009). Setting goals to switch between tasks: effect of cue transparency on children’s cognitive flexibility. *Dev. Psychol.* 45, 782–797. doi: 10.1037/a0015409
- Conway, A., and Stifter, C. A. (2012). Longitudinal antecedents of executive function in preschoolers. *Child Dev.* 83, 1022–1036. doi: 10.1111/j.1467-8624.2012.01756.x
- Costa, A. (2005). “Lexical access in bilingual production,” In *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. De Groot (New York, NY: Oxford University Press), 308–325.
- Costa, A., Hernández, M., Costa-Faidella, J., and Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: now you see it, now you don’t. *Cognition* 113, 135–149. doi: 10.1016/j.cognition.2009.08.001
- Craik, F. I., and Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends Cogn. Sci.* 10, 131–138. doi: 10.1016/j.tics.2006.01.007
- Crivello, C., Kuziyk, 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
- de Bruin, A., Bak, T. H., and Della Sala, S. (2015). Examining the effects of active versus inactive bilingualism on executive control in a carefully matched non-immigrant sample. *J. Mem. Lang.* 85, 15–26. doi: 10.1016/j.jml.2015.07.001
- Diamond, A. (1991). “Neuropsychological insights into the meaning of object concept development,” in *The Epigenesis of Mind: Essays on Biology and Cognition*, eds S. Carey and R. Gelman (Hillsdale, NJ: Erlbaum), 67–110.
- Diamond, A. (2002). “Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy, and biochemistry,” in *Principles of Frontal Lobe Function*, eds D. T. Stuss and R. T. Knight (London: Oxford University Press), 466–504.
- Diamond, A. (2009). All or none hypothesis: a global-default mode that characterizes the brain and mind. *Dev. Psychol.* 45, 130–138. doi: 10.1037/a0014025
- Diamond, A., and Taylor, C. (1996). Development of an aspect of executive control: development of the abilities to remember what I said and to “do as I say, not as I do.” *Dev. Psychol.* 29, 315–334. doi: 10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T

- Donnelly, S., Brooks, P. J., and Homer, B. D. (2015). "Examining the bilingual advantage on conflict resolution tasks: a meta-analysis," in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, eds D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, and P. P. Maglio (Austin, TX: Cognitive Science Society), 596–601.
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited. *Exp. Psychol.* 61, 234–251. doi: 10.1027/1618-3169/a000243
- Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-III*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., Dunn, L. M., and Thériault-Whalen, C. M. (1993). *Echelle de Vocabulaire en Images Peabody: EVIP*. Toronto: Psycan.
- Fay-Stammach, T., Hawes, D. J., and Meredith, P. (2014). Parenting influences on executive function in early childhood: a review. *Child Dev. Perspect.* 8, 258–264. doi: 10.1111/cdep.12095
- Festman, J., and Münte, T. F. (2012). Cognitive control in Russian-German bilinguals. *Front. Psychol.* 3:115. doi: 10.3389/fpsyg.2012.00115
- Festman, J., Rodriguez-Fornells, A., and Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behav. Brain Funct.* 6:5. doi: 10.1186/1744-9081-6-5
- Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., et al. (2010). Bridging language and attention: brain basis of the impact of bilingualism on cognitive control. *Neuroimage* 53, 1272–1278. doi: 10.1016/j.neuroimage.2010.05.078
- 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
- Garraffa, M., Obregon, M., and Sorace, A. (2017). Linguistic and cognitive effects of bilingualism with regional minority languages: a study of Sardinian-Italian adult speakers. *Front. Psychol.* 8:1907. doi: 10.3389/fpsyg.2017.01907
- Gathercole, V. C. M. (2016). "Factors moderating proficiency in bilingual speakers," in *Bilingualism Across the Lifespan: Factors Moderating Language Proficiency*, eds E. Nicoladis and S. Montanari (Washington DC: De Gruyter Mouton and American Psychological Association), 123–140. doi: 10.1037/14939-008
- Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., et al. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Front. Psychol.* 5:11. doi: 10.3389/fpsyg.2014.00011
- Gatt, D., Grech, H., and Dodd, B. (2016). Early lexical expression in children exposed to mixed input: a case of monolingual or bilingual development?. *Int. J. Biling.* 20, 639–665. doi: 10.1177/1367006915572399
- Genesee, F., Nicoladis, E., and Paradis, J. (1995). Language differentiation in early bilingual development. *J. Child Lang.* 22, 611–631. doi: 10.1017/S0305000900009971
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism* 1, 67–81. doi: 10.1017/S1366728998000133
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain Lang.* 36, 3–15. doi: 10.1016/0093-934X(89)90048-5
- Grosjean, F. (2010). *Bilingual*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459
- Hurks, P. P. M., Schrans, D., Meijjs, C., Wassenberg, R., Feron, F. J. M., and Jolles, J. (2010). Developmental changes in semantic verbal fluency: analyses of word productivity as a function of time, clustering, and switching. *Child Neuropsychol.* 16, 366–387. doi: 10.1080/09297041003671184
- Iluz-Cohen, P., and Armon-Lotem, S. (2013). Language proficiency and executive control in bilingual children. *Bilingualism* 16, 884–899. doi: 10.1017/S1366728912000788
- Kang, C., Thoemmes, F., and Lust, B. (2016). Effects of SES on executive attention in Malay-English bilingual children in Singapore. *Bilingualism* 19, 1042–1056. doi: 10.1017/S1366728915000437
- Legacy, J., Zesiger, P., Friend, M., and Poulin-Dubois, D. (2016). Vocabulary size, translation equivalents, and efficiency in word recognition in very young bilinguals. *J. Child Lang.* 43, 760–783. doi: 10.1017/S0305000915000252
- Lengua, L. J., Honorado, E., and Bush, N. R. (2007). Contextual risk and parenting as predictors of effortful control and social competence in preschool children. *J. Appl. Dev. Psychol.* 28, 40–55. doi: 10.1016/j.appdev.2006.10.001
- Lengua, L. J., Moran, L., Zalewski, M., Ruberry, E., Kiff, C., and Thompson, S. (2015). Relations of growth in effortful control to family income, cumulative risk, and adjustment in preschool-age children. *J. Abnorm. Child Psychol.* 43, 705–720. doi: 10.1007/s10802-014-9941-2
- MacSwan, J. (2016). "Code-switching in adulthood," in *Bilingualism Across The Lifespan: Factors Moderating Language Proficiency* eds E. Nicoladis and S. Montanari (Washington, DC: De Gruyter Mouton and American Psychological Association), 163–200.
- Mahone, E. M., and Schneider, H. E. (2012). Assessment of attention in preschoolers. *Neuropsychol. Rev.* 22, 361–383. doi: 10.1007/s11065-012-9217-y
- Matte-Gagne, C., and Bernier, A. (2011). Prospective relations between maternal autonomy support and child executive functioning: investigating the mediating role of child language ability. *J. Exp. Child Psychol.* 110, 611–625. doi: 10.1016/j.jecp.2011.06.006
- Mishina-Mori, S. (2002). Language differentiation of the two languages in early bilingual development: a case study of Japanese/English bilingual children. *IRAL* 40, 211–234. doi: 10.1515/iral.2002.011
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognit. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Morton, J. B., and Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Dev. Sci.* 10, 719–726. doi: 10.1111/j.1467-7687.2007.00623.x
- Namazi, M., and Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. *Int. J. Biling. Edu. Biling.* 13, 597–616. doi: 10.1080/13670050.2010.488288
- Nicoladis, E. (2006). Cross-linguistic transfer in adjective-noun strings by preschool bilingual children. *Bilingualism* 9, 15–32. doi: 10.1017/S136672890500235X
- Nicoladis, E., and Jiang, Z. (2018). Language and cognitive predictors of lexical selection in storytelling for monolingual and sequential bilingual children. *J. Cogn. Dev.* 1–18. doi: 10.1080/15248372.2018.1483370
- Nicoladis, E. and Genesee, F. (1996). A longitudinal study of pragmatic differentiation in young bilingual children. *Lang. Learn.* 46, 439–464. doi: 10.1111/j.1467-1770.1996.tb01243.x
- Nicoladis, E. and Genesee, F. (1997). "The role of parental input and language dominance in bilingual children's code-mixing," in *Proceedings of the 21st Annual Boston University Conference on Language Development*, eds E. Hughes, M. Hughes, and A. Greenhill (Somerville, MA: Cascadia Press), 422–432.
- Paap, K. R., and Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognit. Psychol.* 66, 232–258. doi: 10.1016/j.cogpsych.2012.12.002
- Paap, K. R., Johnson, H. A., and Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences?. *J. Cogn. Psychol.* 26, 615–639. doi: 10.1080/20445911.2014.944914
- 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., and Nicoladis, E. (2007). The influence of dominance and sociolinguistic context on bilingual preschoolers' language choice. *Int. J. Biling. Edu. Biling.* 10, 277–297. doi: 10.2167/beb444.0
- Patterson, J., and Pearson, B. Z. (2004). "Bilingual lexical development: influences, contexts, and processes," in *Bilingual Language Development And Disorders In Spanish-English Speakers*, ed. B. A. Goldstein (Baltimore, MD: Paul H. Brookes), 77–104.
- Peal, E., and Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychol. Monogr. Gen. Appl.* 76, 1–23. doi: 10.1037/h0093840
- Pelham, S. D., and Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 313–325. doi: 10.1037/a0035224



- Plude, D. J., Enns, J. T., and Brodeur, D. (1994). The development of selective attention: a life-span overview. *Acta Psychol.* 86, 227–272. doi: 10.1016/0001-6918(94)90004-3
- Poplack, S. (1985). “Contrasting patterns of code-switching in two communities,” in *Methods V: Papers from the V International Conference on Methods in Dialectology*, ed. H. Warkentyne (Victoria, BC: University of Victoria Press), 363–386.
- Poulin-Dubois, D., Blaye, A., Coutya, J., and Bialystok, E. (2011). The effects of bilingualism on toddlers’ executive functioning. *J. Exp. Child Psychol.* 108, 567–579. doi: 10.1016/j.jecp.2010.10.009
- Prior, A., and MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism* 13, 253–262. doi: 10.1017/S1366728909990526
- Psychology Software Tools, Inc. (2007). *E-Prime Version 2.0*. Available at: <http://www.psnet.com>
- Quay, S., and Montanari, S. (2016). “Early bilingualism: from differentiation to the impact of family language practices,” in *Bilingualism Across the Lifespan: Factors Moderating Language Proficiency*, eds E. Nicoladis and S. Montanari (Washington, DC: De Gruyter Mouton and American Psychological Association), 23–42. doi: 10.1037/14939-003
- Ribot, K. M., and Hoff, E. (2014). “¿Cómo estás?” “I’m good.” Conversational code-switching is related to profiles of expressive and receptive proficiency in Spanish-English bilingual toddlers. *Int. J. Behav. Dev.* 38, 333–341. doi: 10.1177/0165025414533225
- Ricciardelli, L. A. (1992). Bilingualism and cognitive development in relation to threshold theory. *J. Psycholinguist. Res.* 21, 301–316. doi: 10.1007/BF01067515
- Rodriguez-Fornells, A., De Diego Balaguer, R., and Münte, T. F. (2006). Executive functions in bilingual language processing. *Lang. Learn.* 56, 133–190. doi: 10.1111/j.1467-9922.2006.00359.x
- Rodriguez-Fornells, A., van der Lugt, A., Rotte, M., Britti, B., Heinze, H. J., and Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *J. Cogn. Neurosci.* 17, 422–433. doi: 10.1162/0898929053279559
- Rubin, O., and Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1477–1491. doi: 10.1037/0278-7393.31.6.1477
- Ruff, R. M., Light, R. H., Parker, S. B., and Levin, H. S. (1997). The psychological construct of word fluency. *Brain Lang.* 57, 394–405. doi: 10.1006/brln.1997.1755
- Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., and Boyce, W. T. (2011). Family socioeconomic status and child executive functions: the role of language, home environment, and single parenthood. *J. Int. Neuropsychol. Soc.* 17, 120–132. doi: 10.1017/S1355617710001335
- Sergeant, J. A., Geurts, H., and Oosterlaan, J. (2002). How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? *Behav. Brain Res.* 130, 3–28. doi: 10.1016/S0166-4328(01)00430-2
- Shao, Z., Janse, E., Visser, K., and Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Front. Psychol.* 5:772. doi: 10.3389/fpsyg.2014.00772
- Silva-Corvalán, C., and Treffers-Daller, J. (eds). (2016). *Operationalising and Measuring Language Dominance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107375345
- Sioufi, R., Bourhis, R. Y., and Allard, R. (2016). Vitality and ethnolinguistic attitudes of acadians, franco-ontarians and francophone quebecers: two or three solitudes in Canada’s bilingual belt? *J. Multiling. Multicult. Dev.* 37, 385–401. doi: 10.1080/01434632.2015.1072205
- Sorge, G. B., Toplak, M. E., and Bialystok, E. (2017). Interactions between levels of attention ability and levels of bilingualism in children’s executive functioning. *Dev. Sci.* 20:e12408. doi: 10.1111/desc.12408
- Soveri, A., Rodriguez-Fornells, A., and Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Front. Psychol.* 2:183. doi: 10.3389/fpsyg.2011.00183
- Tao, L., Marzecová, A., Taft, M., Asanowicz, D., and Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: the role of age of acquisition. *Front. Psychol.* 2:123. doi: 10.3389/fpsyg.2011.00123
- Tao, L., Taft, M., and Gollan, T. H. (2015). The bilingual switching advantage: sometimes related to bilingual proficiency, sometimes not. *J. Int. Neuropsychol. Soc.* 21, 531–544. doi: 10.1017/S1355617715000521
- Tare, M., and Linck, J. (2011). “Exploring bilingual cognitive advantages when controlling for background variables,” in *52nd Annual Meeting of the Psychonomic Society*, Seattle, WA.
- Thomas-Sunesson, D., Hakuta, K., and Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *Int. J. Biling. Edu. Biling.* 21, 197–206. doi: 10.1080/13670050.2016.1148114
- Troyer, A. K., Moscovitch, M., and Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology* 11, 138–146. doi: 10.1037/0894-4105.11.1.138
- Valian, V. (2015). Bilingualism and cognition. *Bilingualism* 18, 3–24. doi: 10.1017/S1366728914000522
- von Bastian, C. C., Souza, A. S., and Gade, M. (2016). No evidence for bilingual cognitive advantages: a test of four hypotheses. *J. Exp. Psychol. Gen.* 145, 246–258. doi: 10.1037/xge0000120
- Weckerly, J., Wulfeck, B., and Reilly, J. (2001). Verbal fluency deficits in children with specific language impairment: slow rapid naming or slow to name? *Child Neuropsychol.* 7, 142–152. doi: 10.1076/chin.7.3.142.8741
- Yow, W. Q., and Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: why variations in bilingual experiences matter. *Front. Psychol.* 6:164. doi: 10.3389/fpsyg.2015.00164
- Yow, W. Q., Tan, J. S., and Flynn, S. (2017). Code-switching as a marker of linguistic competence in bilingual children. *Biling. Lang. Cogn.* 227–258. doi: 10.1017/S1366728917000335

**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 Nicoladis, Hui and Wiebe. 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.



# Language Dominance Affects Bilingual Performance and Processing Outcomes in Adulthood

Eloi Puig-Mayenco<sup>1\*</sup>, Ian Cunnings<sup>1</sup>, Fatih Bayram<sup>2</sup>, David Miller<sup>1</sup>, Susagna Tubau<sup>3</sup> and Jason Rothman<sup>1,2</sup>

<sup>1</sup> School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom, <sup>2</sup> Department of Language and Culture, UiT the Arctic University of Norway, Tromsø, Norway, <sup>3</sup> Departament de Filologia Anglesa i Germanística, Universitat Autònoma de Barcelona, Barcelona, Spain

## OPEN ACCESS

### Edited by:

Esther Rinke,  
Goethe-Universität Frankfurt am Main,  
Germany

### Reviewed by:

Maria Carmen Parafita Couto,  
Leiden University, Netherlands  
Mike Putnam,  
Pennsylvania State University,  
United States

### \*Correspondence:

Eloi Puig-Mayenco  
e.puig-mayenco@pgr.reading.ac.uk

### Specialty section:

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

**Received:** 27 March 2018

**Accepted:** 21 June 2018

**Published:** 26 July 2018

### Citation:

Puig-Mayenco E, Cunnings I,  
Bayram F, Miller D, Tubau S and  
Rothman J (2018) Language  
Dominance Affects Bilingual  
Performance and Processing  
Outcomes in Adulthood.  
Front. Psychol. 9:1199.  
doi: 10.3389/fpsyg.2018.01199

This study examines the role of language dominance (LD) on linguistic competence outcomes in two types of early bilinguals: (i) child L2 learners of Catalan (L1 Spanish-L2 Catalan and, (ii) child Spanish L2 learners (L1 Catalan-L2 Spanish). Most child L2 studies typically focus on the development of the languages during childhood and either focus on L1 development or L2 development. Typically, these child L2 learners are immersed in the second language. We capitalize on the unique situation in Catalonia, testing the Spanish and Catalan of both sets of bilinguals, where dominance in either Spanish or Catalan is possible. We examine the co-occurrence of Sentential Negation (SN) with a Negative Concord Item (NCI) in pre-verbal position (Catalan only) and Differential Object Marking (DOM) (Spanish only). The results show that remaining dominant in the L1 contributes to the maintenance of target-like behavior in the language.

**Keywords:** language dominance, Negative Concord Items, Differential Object Marking, early bilinguals, Catalan/Spanish

## INTRODUCTION

A large body of studies involving early childhood bilinguals examine the development of linguistic competence during the acquisition process itself, often focusing on how bilingual acquisition is qualitatively similar or different to monolinguals during the developmental period of language learning (see Meisel, 2011; Serratrice, 2013; Nicoladis, 2018 for a review). Furthermore, studies concerned with adult second language acquisition or first language attrition largely focus on similar processes; however, they do so with inherently different contexts concerning age of onset and other deterministic variables (see Rothman and Slabakova, 2017; White, 2018; Wulff and Ellis, 2018; Yilmaz and Schmid, 2018 for updated reviews from various paradigmatic approaches). The focus, thus, is on the acquisition of another language starting in adulthood and the ensuing developmental consequences, as in the case of attrition, on the maintenance of previously acquired languages.

A notable exception to the trends in the above literature is the work on heritage speaker (HS) bilingualism (see Montrul, 2008, 2016; Rothman, 2009; Benmamoun et al., 2013; Kupisch and Rothman, 2016; Polinsky, 2018). To date, the focus within HS bilingualism has been to examine adult steady-state grammars of (at least) the minority (heritage) language acquired in early childhood. The heritage language is one of the HS's L1s, either acquired simultaneously with the societal majority language (2L1) or as the unique L1 in the case of child L2 acquisition whereby immigration occurs before or at school age (roughly 5–6 years old). Thus, HSs are a subtype of

native speaker (Rothman and Treffers-Daller, 2014). This is interesting given that studies generally reveal that adult HS grammars reflect both dominance in the majority language (i.e., whether a simultaneous L1 or child L2 for the HS) and degrees of non-monolingual-like variability in the heritage L1 (see Montrul, 2008, 2016; Benmamoun et al., 2013).

The typical HS outcomes are, at first glance, surprising in light of child 2L1 and child L2 studies that generally demonstrate greater conformity, whether in qualitative similarities in development and/or ultimate attainment (see for review Meisel, 2011; Haznedar, 2013; Chondrogianni, 2018). After all, HSs tested as adults are the outcomes of 2L1 or child L2 acquisition. As such, we are left to wonder why they differ to such a degree in adulthood from the seemingly successful trajectory that research on child bilingualism suggests they were on (Kupisch and Rothman, 2016). In recent years, several researchers have suggested that HSs' grammatical outcomes in adulthood likely highlight distinctive acquisition paths, reflecting the individual realities of personal, minority language/bilingual situations for variables that become more deterministic in later childhood (e.g., Putnam and Sánchez, 2013; Kupisch and Rothman, 2016). In other words, in addition to effects of L1 attrition and/or arrested development at the individual HS level, linguistic and extra-linguistic variables conspire to change the path of HS grammatical development and, thus, explain the default trend of considerable variation both between HSs and monolinguals, as well as other HSs. The emerging literature has highlighted the following variables, among others: (1) the quality of input affected by language contact (L1 attrition of the older generation); (2) the lack of literacy in the heritage language; (3) the influence of formal properties (features) of the majority language, altering the formal HS learning task; and (4) being outside a bilingual community representing true diglossia. All of these variables reduce opportunities to use the minority language and receive/uptake (quantity/quality) input (e.g., Putnam and Sánchez, 2013; Kupisch and Rothman, 2016; Bayram et al., 2017; Karayayla and Schmid, 2017; Karayayla, 2018).

In the vast majority of work on heritage bilingualism to date, the default context is one of a distinct majority language that subsumes the minority one in all aspects of societal distribution (e.g., only the heritage community is bilingual in the languages under investigation whereby education is typically in the majority language) and there is a palpable imbalance of prestige between the two languages. It is this situation itself that promotes the abovementioned imbalance in extra-linguistic variables. If the unequal distribution of these extra-linguistic variables across various HS groups or individuals factors into the unique outcomes of HSs (Lloyd-Smith et al., submitted), then we should see monolingual-to-bilingual differences significantly diminish or be eradicated in the adult outcomes of 2L1 speakers and especially child L2 bilinguals when the context for bilingualism is more favorable. This should be especially true when the society itself is bilingual in the same languages.

The case of Catalonia is an ideal environment to put the above to test as successful bilingualism is the default in this setting, inclusive of the purposeful efforts in place in the education system to ensure that all young people are formally literate and

educated in both languages. The fact that there is near universal success in Catalan-Spanish bilingual outcomes does not negate the fact that the order of acquisition of both languages can vary across individuals, and that depending on where in Catalonia one grows up it could be said that one or the other is more dominant. Moreover, successful bilingualism at the community level does not preclude cross-linguistic influence in developing bilingual grammars. Looking at how differences might obtain even in such a context, and whether this correlates/varies with order of acquisition and other measures of relevant dominance (patterns of use) in one or the other language, can augment the heritage speaker literature more generally. Minimally, showing what is similar and distinct both between our bilinguals here and more typical HS outcomes can reveal what is likely to differ between monolingual and child bilingual outcomes in adulthood universally vs. what obtains independently as the byproduct of the less-than-ideal bilingual environments HSs tend to grow up in.<sup>1</sup>

In the present study, all bilinguals are formally trained in literacy in both languages. We provide data from two groups of Catalan-Spanish bilinguals who were born and raised in Osona, Catalonia where dominance in Catalan is the default.<sup>2</sup> The first group comprises child L2 learners of Spanish (L1-Cat-L2-Sp) and the second group comprises child L2 learners of Catalan (L1-Sp-L2-Cat). The present study is also one of a select few that tests each bilingual group in both languages, which is needed to understand more fully how the languages of a bilingual interact

<sup>1</sup>Following neutral definitions for HS inclusion such as (Rothman, 2009) where deficit outcomes are not part of the criteria of defining factors of HSs, one could convincingly make an argument that children of parents who immigrate from monolingual Southern Spain and raise their children even in rural Catalonia as (virtual) monolingual Spanish speakers until they go to school are indeed a specific subtype of HSs. Of course, they would be exposed tangentially to Catalan, just like Spanish HSs are to English in the ubiquitously studied case of Spanish HS bilingualism in the US, before schooling starts. Not being significantly exposed to the other (societal majority language) is in fact even more possible in a place like Catalonia where everyone is bilingual, such that each individual a child encounters can effortlessly switch to the language the child prefers. Indeed, the societal status of the languages and all this entails for use and exposure differ significantly between the two contexts we are comparing, but these differences are exactly what we capitalize on as they permit a teasing out of variables otherwise not possible. Whether or not such speakers are in fact accepted as HSs by all is not important for the purposes of our argumentation. In the case they are not, we can only hope that the reason does obtain because of the general success of Catalan-Spanish bilingualism itself. HS bilingualism does not necessarily entail lack of success (Kupisch and Rothman, 2016), which is immediately clear if one accepts that our Spanish L1-Child L2 Catalan group should count as a subtype of HSs.

<sup>2</sup>Language dominance and its measurement have been widely debated (e.g., Bialystok, 2007; Montrul, 2015; Schmeißer et al., 2015; Silva-Corvalán and Treffers-Daller, 2015; Unsworth, 2015). In this study, by dominance we refer to patterns of preferred use and usage frequency in daily life following Unsworth's (2015) suggestion that language exposure/use patterns might be taken as a proxy of LD. As such one should not infer anything with regard to proficiency per se. As stated and will be quantified below, all participants are highly proficient, performing on the standardized measures for both languages with no statistical difference. Following Montrul (2015) and Schmeißer et al. (2015), we do not assume a direct relationship between dominance and proficiency per se, even if in unbalanced bilingual environments there tends to be correlations. As discussed in Perpiñán (2017) and quantified below, Catalan-Spanish bilinguals tend to be highly proficient in both languages, thus, correlating proficiency and language dominance in this context of balanced bilingualism might not prove useful.

and how this might differ across bilingual groups depending on factors such as the ones that differentiate our bilingual groups from those pertaining to typical HS environments.

Given this relatively unique environment, one can find bilinguals who are more dominant in one or the other language while highly proficient and literate in both. It is not uncommon to find a child L2 learner of Catalan in Catalonia who remains dominant in their L1 (Spanish), unlike the typical case of immersed child L2 learners. What is especially interesting about Osona is that the minority (Spanish) and the majority (Catalan) languages of the immediate regional society, which should matter most, are the opposite in the national context. This variable will be considered pertaining to the generalizability of the results.<sup>3</sup> However, Catalonia is certainly not the only context in the world where this applies. Beyond contributing to the literature by offering a study that examines somewhat different conditions for the outcomes of a case of child L2 bilingualism in adulthood (as well as potential consequences to their L1), we endeavor to show how capitalizing on the unique positioning of variables that contexts like Catalonia present by default can inform important questions of theoretical relevance. Minimally, isolating some of these extra-linguistic variables has the potential to explain individual variation across bilingual speakers of the same two languages, even when both languages are readily available in the environment and supported via education.

We investigate two subtle phenomena in Spanish and Catalan: (1) the co-occurrence of Sentential Negation (SN) with a Negative Concord Item (NCI) in pre-verbal position, allowed in Catalan yet disallowed in Spanish and (2) Differential Object Marking (DOM), obligatory in Spanish but not part of the Catalan grammar. We chose these phenomena because they are claimed to be sensitive to variation in the adult grammars of childhood bilinguals (Montrul, 2004; Déprez et al., 2015) in other contexts.

## THEORETICAL BACKGROUND

Our chosen properties are of particular interest because they allow us to look at whether order of acquisition and language dominance play a role in the expansion of the distribution of a specific linguistic domain. Negative Concord Items (NCIs) in Catalan have a wider distribution [with and without sentential negation (SN)] than in Spanish (without SN). The distribution of DOM in Catalan and Spanish also presents differently, whereby Spanish has a wider distribution of DOM than Catalan. Though variable across dialects, DOM in Spanish *par excellence* (i.e., across dialects) is obligatory in certain cases, such as marking accusative [+animate/+specific] objects. Indeed DOM is subject to semantic and discourse constraints in particular contexts

(e.g., as it interfaces with modality, indicative vs. subjunctive in embedded clauses); however, in the domain of DOM we focus on there is no such considerations affecting its use. In other words, it is a morphosyntactic reflex of obligatory (accusative) case marking. DOM is more restricted in Catalan and is ungrammatical in the Spanish-canonical position of [+animate/+specific] objects in their base-generated position. In both cases, the smaller distribution is subsumed by the language with the larger distribution: (a) all contexts in which DOM exists in Catalan exist in Spanish, but Spanish has more obligatory DOM contexts and (b) all contexts where Spanish NCI is allowed hold true for Catalan, although Catalan also allows it with SN. And so, assuming that influence will proceed from a subset to a superset, choosing these two domains allows us to look without prejudice for one language over the other into whether CLI will obtain accordingly in relatively balanced bilingualism (no differences related to relative dominance), or if CLI is conditioned by relative dominance in one or the other language.

## Negative Concord Items (NCIs) in Catalan and Spanish

NCIs have been argued to be negative Universal Quantifiers (Haegeman and Zanuttini, 1991; Zanuttini, 1991), positive Polarity Items (Laka, 1990), negative indefinites (Suñer, 1995), and non-negative indefinites (Zeilstra, 2004; Tubau, 2008). Herein, we adopt Zeilstra's (2004) analysis of NCIs specifically for Catalan and Spanish while considering some modifications offered by Espinal and Tubau (2016).

Both Catalan and Spanish are Negative Concord (NC) languages. NC languages are typified by two main varieties: strict NC Languages, in which the sentential negation (SN) is always obligatory, as in Romanian; and non-strict NC languages, in which the sentential negation is obligatory when the NCI is in post-verbal position and disallowed when the NCI is in pre-verbal position, such as in Spanish. Note that there is a third option that is universally marked, which is essentially a weak version of the strict NC language option described above. In such cases, the negative marker is possible with a pre-verbal NCI but not obligatory. Among the members of the Romance family, Catalan seems to be the only language that allows for optionality of the negative marker when the NCI is in pre-verbal position (Quer, 1993; Vallduví, 1994; Espinal, 2000; Tubau, 2008). All of this can be seen in the grammaticality of (1a-1b) and (4a-4b), the ungrammaticality in (2a-2c) and the variation in grammaticality of (3a-3b) and (4a-4b).

- |     |    |         |                                 |             |              |              |
|-----|----|---------|---------------------------------|-------------|--------------|--------------|
| (1) | a. | No      | vindrà                          | ningú       | a la festa   | Catalan      |
|     | b. | No      | vendrà                          | nadie       | a la fiesta  | Spanish      |
|     |    |         | Not                             | will.come   | n-person     | to the party |
|     |    |         | "Nobody will come to the party" |             |              |              |
| (2) | a. | *Vindrà | ningú                           | a la festa  | Catalan      |              |
|     | b. | *Vendrà | nadie                           | a la fiesta | Spanish      |              |
|     |    |         | Will.come                       | n-person    | to the party |              |
|     |    |         | "Nobody will come to the party" |             |              |              |

<sup>3</sup>Although the majority of people in Catalonia are bilinguals speaking both languages (99% of the population speak and understand Spanish and 96.5% of the population speak and understand Catalan, Idescat, 2013). In this study, we targeted an area where Catalan is clearly the majority dominant language of the environment (73% of Catalan in the daily use in this area as opposed to 43% in Catalonia as a whole, Idescat, 2013). The reader is referred to Illamola (2015) for an in-depth presentation of the sociolinguistic patterns and the use of both Catalan and Spanish in the specific town (Manlleu, Osona) where the data were collected.



- (3) a. Ningú no vindrà a la festa *Catalan*  
 b. \*Nadie no vendrá a la fiesta *Spanish*  
 n-person not Will.come to the party  
 “Nobody will come to the party”
- (4) a. Ningú vindrà a la festa *Catalan*  
 b. Nadie vendrá a la fiesta *Spanish*  
 n-person Will.come to the party  
 “Nobody will come to the party”

## Differential Object Marking

DOM is the overt morphological expression used by some languages to mark Case on some accusative objects. Spanish is known to be a DOM language (e.g., Leonetti, 2004; López, 2012). Unlike Spanish, Catalan presents a less clear case;<sup>4</sup> however, it is well attested that in both Standard Catalan and the Central Catalan dialect, which are the dialects relevant to our bilingual groups herein, DOM is not expected (Escandell-Vidal, 2009; GIEC, 2016).

Rodríguez-Mondoñedo (2007) suggests that there are two important dimensions which help determine the marking of the object: animacy and specificity.<sup>5</sup> As pointed out by Leonetti (2004), animacy has been labeled as the dominant factor. If we use these two dimensions, there are four possible scenarios for objects: [+specific/+animate], [+animate/−specific], [−animate/+specific] and [−animate, −specific]. In Spanish, the object is obligatorily marked when the object is [+specific, +animate] as in (5a-b).

- (5) a. María vio a Laura *Spanish*  
 b. \*María vio ∅ Laura  
 María saw DOM Laura  
 “María saw Laura”

When the object is [−animate/+specific] or [−animate/−specific], then the object is obligatorily unmarked. The case of [+animate/−specific] can be marked, this this depends on various semantic and discourse features that we highlight here for the sake of being complete. As we only focus on [+animate/+specific] contexts in which the marker is obligatory and, to our knowledge, not subject to dialectal variation as other subtypes are, we will not comment further on the inherent variation of DOM cross-dialectically.

Importantly, the distribution of DOM in Standard Catalan and Central Catalan is more restricted than in Spanish. For example, in Catalan [+animate/+specific] full DP objects are left unmarked (compare 8a-b). Thus, the experiments herein contain full DPs.

- (6) a. La Maria buscava ∅ la Laura *Catalan*  
 \*La Maria buscava a la Laura  
 b. The Maria looked.for DOM the Laura  
 ‘Maria was looking for Laura’

However, the fact that the marker does not appear in this context does not mean that DOM is non-existent in Catalan, a point to which we return in the discussion when we discuss the input. As reported in (GIEC, 2016), DOM is required when the [+animate/+specific] is a full pronoun or in cases where the full DP object is found in focalized constructions. In sum, the above illustrates that DOM occurs in certain contexts in Catalan, but, crucially, does not occur in the context under investigation, which entails that its distribution is somewhat more restricted than in Spanish.

## Studies on Catalan-Spanish Bilingualism

Although there is a line of research that has looked at developmental patterns in Catalan-Spanish bilinguals (e.g., Bel, 1996, 2001, 2003; Bosch and Sebastian-Gallés, 2001; Guijarro-Fuentes and Marinis, 2009; Simonet, 2011, 2014; Guijarro-Fuentes, 2012; Illamola, 2015; Perpiñán, 2017), there are relatively few studies that have examined Catalan/Spanish bilingualism outcomes in adulthood. In that respect, Perpiñán (2017) stands out as a noteworthy study examining the effects of early bilingualism in adulthood in the domain of non-personal clitics in Catalan that are lacking in Spanish [i.e., the partitive clitic (*en*) and the locative clitic (*hi*)]. Her results show that the group of Spanish-dominant speakers were significantly less sensitive to instances of ungrammaticality than the Catalan-dominant speakers. This is an expected, though significant, result. It is often the case that bilingual knowledge differs significantly from the anticipated monolingual outcome. However, in a context like Catalonia where relatively balanced bilingualism is likely, and both languages are supported at all levels, it is reasonable to hypothesize that bilingual grammars would differ less from monolinguals than in other cases of bilingualism. Indeed, this expectation has some evidence. Recall that the Catalan-dominant group is also bilingual, yet conforms to monolingual norms significantly more and thus reveals that dominance, even in a society where access to both languages is ubiquitous, matters.

Studies like Perpiñán (2017) are significant because they show more of the same, that is, they highlight the effects of bilingualism that exist despite a context that is maximally supportive for success and, crucially, seem to suggest that dominance—and not proficiency *per se*—matters. Consequently, it is clear that bilingualism effects are real, meaning some differences in bilingual grammars obtain because of bilingualism itself (Sorace, 2011) and not merely because of extra-linguistic considerations such as poor access to input, low prestige of a weaker language, etc., that define the reality of many, or perhaps most, of the realities of individual bilinguals. However, the fact that bilingualism itself, even under ideal contexts, can invite monolingual base-line differences—bilingualism is not multiple instances of monolingualism in the same mind/brain (Grosjean, 1989)—does not mean that lack of such a supportive

<sup>4</sup>DOM is present in some varieties of Catalan (potentially stemming from cross-linguistic influence issues related to Spanish as well), e.g., Balearic and Valencian Catalan (Escandell-Vidal, 2009; GIEC, 2016), however, DOM is definitively not part of the dialects spoken by the participants included in our study.

<sup>5</sup>The notion of specificity has been widely debated in the literature. We take a widely accepted notion that specificity expresses a semantic property of the element that determines the referent of the element in a particular way (see Farkas, 1995; von Stechow, 2002; von Stechow and Kaiser, 2003; Leonetti, 2004; López, 2012; for a more detailed description and analysis of specificity).

environment and the entailed beneficial byproducts of it would not further exaggerate monolingual-bilingual differences. In other words, what would the speakers in Perpiñán (2017) look like if they grew up in a less supportive bilingual environment, such as a typical HS environment? On the basis of this work, we expect some cross-linguistic influences in our Catalan-Spanish bilinguals, but, like Perpiñán (2017), we expect them to be subtle differences and not subject to a large amount of inter-speaker variation as is the default when typical HSs are tested.

## DOM and NCIs in Catalan-Spanish Bilingualism

Although there has been considerable work in recent years examining the acquisition of DOM in L2 Spanish (e.g., Farley and McCollam, 2004; Montrul, 2004; Bowles and Montrul, 2008; Guijarro-Fuentes and Marinis, 2009; Guijarro-Fuentes, 2012), and how it appears in HS Spanish grammars in North America (e.g., Montrul, 2004; Montrul and Bowles, 2009; Montrul and Sánchez Walker, 2013) to various degrees of successful convergence, we are aware of only one study that examines it in the context of Catalan-Spanish bilingualism (Guijarro-Fuentes and Marinis, 2009). In this study, the authors showed that Catalan-Spanish sequential bilinguals, although outperforming English learners of L2 Spanish, were still considerably different from Spanish monolinguals in the sense that they over-accepted the accusative makers in contexts where they were not grammatical. Guijarro-Fuentes and Marinis (2009) make no mention of having tested for language dominance (LD); however, given the context and the fact that they are home speakers of Catalan, it is fair to assume that if they were not balanced bilinguals, LD for the group would be in Catalan. From their results, we know that grammatical sensitivity to Spanish DOM can be affected by the more restricted domain of DOM in Catalan. What we do not know, however, is if the restricted DOM in Catalan can be affected (expanded) by Spanish in the opposite direction of LD. This latter point is addressed by the present bi-directional study.

Contrary to the case of DOM, there is a dearth of available studies looking at NCIs from an acquisition perspective in the Catalan-Spanish bilingual literature. However, experimental research in syntax has been done to corroborate current theoretical descriptions in both languages. Déprez et al. (2015) examine the interpretation of pre-verbal NCI when occurring with Sentential Negation (SN) in Catalan. They examined whether the co-occurrence of the SN would trigger Double Negation (DN) readings of the NCI as opposed to NC readings. Their findings suggest that the default reading of a pre-verbal NCI in Catalan with the SN is a generally an NC one, which is not possible in Spanish (Déprez et al., 2015; Espinal et al., 2016).

## RESEARCH QUESTIONS AND PREDICTIONS

The main overarching research question that motivated the present study was:

- a. What role do order of acquisition and language dominance have—independent of overall linguistic proficiency—in the competence and performance of early child bilinguals tested in adulthood?

As is true of all specific research, overarching questions must be packaged in testable ways, examining specific domains of grammar in specific sub-groups of participants under appropriate contexts as proxies. And so, question (a) can be asked as (b):

- b. What is the respective role of order of acquisition and dominance in Catalan and Spanish regarding the competence and performance outcomes of NCIs and DOM among early child bilinguals tested in adulthood?

Our hypotheses are:

- c. Language dominance matters. Cross-linguistic influence (CLI) from Catalan-to-Spanish and Spanish-to-Catalan is *a priori* possible for both groups. Perhaps, irrespective of dominance, some CLI will be noted. We also predict that greater degrees of CLI might correlate to relative dominance, in which case there would be significant differences across the two groups. We also hypothesize that the domain of grammar matters. CLI is conditioned by the comparative status of the properties in the two grammars; CLI will influence expansion in the grammar with a more restricted distribution. This means we expect emerging optionality in Spanish NCIs and/or expansion of DOM in Catalan contexts where it is prohibited via influence of the larger distribution in the other grammar, but not vice versa. That is, Catalan may lose optionality in NCI interpretation or Spanish may lose DOM in canonical contexts not supported by Catalan. We further predict that there could be differences across the two domains of grammar, whereby NCIs are either not affected or they are less affected because Catalan's larger grammar reflects optionality which contains the Spanish obligatory option, compared to the case of DOM where Spanish, the larger grammar, reflects obligatory use of DOM in unattested contexts of Catalan.

## METHODOLOGY

### Participants

We tested two groups of participants who differ in their order of acquisition and their reported language use and exposure. We included only participants whose proxy for dominance, assessed by means of reported use and exposure via the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007), indicated accordance between their L1 (Spanish or Catalan) and their dominance in adulthood. Although the default assumption of HS bilingualism in general is that dominance in adulthood will be in the L2, we are interested in knowing what effects bilingualism has in the case that one can and does remain dominant in their L1 even if, like the typical HS situation, it is not the preferred, majority language of the bilingual situation. Thus, in an effort to not muddy the waters, we examined bilinguals who were balanced in proficiency across the two languages,

yet each group remains dominant in their L1. The first group of participants consists of Spanish-Catalan bilinguals who were exposed to Spanish from birth and Catalan at schooling age: L1Sp-L2Cat speakers ( $n = 23$ ). Though the schooling system is generally in Catalan and the language of the environment is Catalan, they reported high levels of use of and exposure to Spanish.<sup>6</sup> The second group is comprised of Catalan-Spanish bilinguals who were exposed to Catalan at home and whose first significant exposure to Spanish was at school age: L1Cat-L2Sp speakers ( $n = 21$ ).

All participants were vetted to ensure fulfillment of the inclusion criteria: (1) Catalan/Spanish bilinguals with no other native languages, (2) minimum proficiency in any foreign languages,<sup>7</sup> (3) high native scores in both Catalan and Spanish proficiency tests, and (4) residence in the geographical (Osona) area where data were collected (Central Catalan dialect). Spanish proficiency was measured through the DELE, which is standardly used as a measure of proficiency in the field (e.g., Montrul and Slabakova, 2003; Slabakova and Montrul, 2003; Bruhn de Garavito and Valenzuela, 2008; Slabakova et al., 2012). Catalan proficiency was measured using a part of the *Certificat Superior de Llengua Catalana* implemented by the *Centre de Normalització Lingüística*.

The Leap-Q was used to assess overall language use and exposure, which we used as a proxy for dominance. We also examined answers from the Catalan version of the Leap-Q questionnaire. We first looked at their responses of the questionnaire:<sup>8</sup> question 1 (dominant language), question 3 (exposure to each language), question 5 (use of both languages); and their responses in the questions for each language: question 2, 4, and 5 (exposure in different environments). Such questions probed self-reported percentages of use and exposure to each language, as well as assessing amount of exposure on a scale from 1 to 10 (1 = not much exposure; 10 = a lot of exposure). A participant was categorized as dominant in one language or the other when two or three of the following conditions were met: (a) reported exposure in one language was higher than the other, (b) reported use in one language was higher than the other, and (c) the self-rated exposure to one of the language was higher than the other. **Table 1** provides the participant profiles after the inclusion criteria had been applied.

This study was carried out in accordance with the recommendations of Research Ethics Committee. The protocol was approved by the School of Psychology and Clinical

**TABLE 1** | Details of the participants.

	L1Sp-L2Cat ( $N = 23$ )	L1Cat-L2Sp ( $N = 21$ )
Mean age	22	20
Proficiency in Catalan	34/40	36/40
Proficiency in Spanish	46/50	45/50
Dominant Language	Spanish	Catalan
Mean (%) exposure to Spanish	59%	18%
Mean (%) exposure to Catalan	41%	82%
Mean (%) use of Spanish	66%	85%
Mean (%) use of Catalan	34%	15%
Rate (1-to-10) of exposure to Spanish	6.5/10	2.5/10
Rate (1-to-10) of exposure to Catalan	3.5/10	7.5/10

Language Science's Research Ethics Committee at the University of Reading. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## Tasks

Participants took part in two separate experimental tasks: an off-line Grammaticality Judgement Task and a non-cumulative, moving window Self-Paced Reading (SPR) Task in both languages.<sup>9</sup> Presentation by language was counter-balanced: half of the participants were asked to do the Catalan experiments first and vice versa. All the tasks were delivered using IBEX FARM software and the experiments were done in a controlled lab environment.

### Grammaticality Judgement Task

All participants completed two Grammaticality Judgement Tasks: one in Spanish and another in Catalan. Each task consisted of 48 items which were distributed across six conditions (four target conditions + two filler conditions) with eight items per condition. The four target conditions are described below.

Condition (a) (NCI+SN) consisted of sentences with a NCI [*nadie*, Sp or *ningú*, Cat; "*nobody*" in pre-verbal position followed by the negative marker *no*. This structure is ungrammatical in Spanish, but grammatical in Catalan [see examples in (7a-b)]. The items in condition (b) (NCI-SN) were target sentences containing an NCI without the negative marker—a structure which is acceptable in both Catalan and Spanish. [See examples in (8a-b)].

Condition (c) (+DOM) consisted of items with a [+animate, +specific] marked DP object by the Accusative Marker "*a*." In Spanish, this is grammatical, whereas in Central Catalan and Standard Catalan it is ungrammatical. Condition d) (−DOM) consisted of items in [+animate, + specific] without the accusative marker. This is grammatical in Catalan and ungrammatical in Spanish. See examples (9a-b) and (10a-b).

The sentences in these two tasks were judged on a 6-point Likert scale where 1 was completely odd and 6 was completely

<sup>6</sup>It is crucial to recall here that our claim of the clear-cut nature of Catalan being the majority language relates to the specific location, central rural Catalonia. Such a claim would not be so evidently true in large Metropolitan areas, such as Barcelona. For example, Sorolla (2009) shows that children whose native language (home language) is Spanish tend not to use Catalan as much, even with peers, despite being educated in Catalan primarily. This is not so surprising given the demographics of such a large metropolitan area, the very reason we decided to test in rural Catalonia where the incidence of ethnic Catalans as discussed in footnote 3 (see work by Illamola, 2015).

<sup>7</sup>We would have wanted to exclude participants with knowledge of foreign languages, but English is mandatory in the system and they all had, at least, minimal exposure to it.

<sup>8</sup>See the corresponding questions in the English version of the Leap-Q questionnaire.

<sup>9</sup>The raw data of the experimental task will be made available by the authors, without undue reservation, to any qualified researcher.

- (7) a. Ningú no portarà globus per la festa de demà *Catalan*  
 b. \*Nadie no traerá globos para la fiesta de mañana *Spanish*  
 Nobody not will.bring balloons for the party of tomorrow  
 ‘Nobody will bring balloons for tomorrow’s party’
- (8) a. Ningú portarà globus per la festa de demà *Catalan*  
 b. Nadie traerá globos para la fiesta de mañana *Spanish*  
 Nobody will.bring balloons to the party of tomorrow  
 ‘Nobody will bring balloons for tomorrow’s party’
- (9) a. \*Les noies coneixeran a la Maria a la festa de demà *Catalan*  
 b. Las chicas conocerán a María en la fiesta de mañana *Spanish*  
 The girls will meet DOM Mary in the party of tomorrow  
 “The girls will meet Mary in tomorrow’s party”
- (10) a. Les noies coneixeran la Maria a la festa de demà *Catalan*  
 b. \*Las chicas conocerán María en la fiesta de mañana *Spanish*  
 The girls will meet Mary in the party of tomorrow  
 “The girls will meet Mary in tomorrow’s party”

natural. There was also an option of “I’m not sure.” Participants were instructed to answer as fast as possible and to leave aside any prescriptive judgements by rating the sentences according to their own intuitions. There were eight practice items, after which the experimental items started.

### Self-Paced Reading Task

The Self-Paced Reading Task was also administered in each language and used the same four experimental conditions [(a) NCI+SN, (b) NCI-SN, (c) +DOM and (d) -DOM], each of which contained eight items, in addition to four filler conditions ( $n = 64$ ). The filler conditions consisted of sentences with similar structures but without the occurrence of NCIs or DOM. Each item was divided into regions of interest which were then used to examine reaction times and spill-over effects. An example of this division can be seen in (11a-b) below

- (11) a. Ningú / no / portarà / globus / per / la festa / de / demà *Catalan*  
 b. Nadie / no / traerá / globos / para / la fiesta / de / mañana *Spanish*

We created two lexical sub-contexts such that the sentences did not become repetitive: half of the experimental items had vocabulary related to a party and half of them to a market. Examples of experimental items can be seen in (7–10) above as we used similar sentences to those in the GJT. Participants were instructed to read the sentences at a normal pace and respond to comprehension questions. They were instructed to do the first three items and to ask any questions, after which the experiment started with six distractor items, then the 64 items were presented in a random fashion.

## RESULTS

### Grammaticality Judgement Tasks

#### Descriptive Analysis

Table 2 presents the Grammaticality Judgement data in both Catalan and Spanish from the two experimental groups: L1Cat-L2Sp ( $n = 21$ ) and L1Sp-L2Cat ( $n = 23$ ) for the two properties and all the conditions. In order to conduct the statistical analysis, the responses in the 6-point Likert scale were converted using a

binary coding: responses from 1 to 3 were coded as rejection “0” and responses from 4 to 6 were coded as acceptance “1”<sup>10</sup>

The results for the NCI+SN and NCI-SN conditions in Table 2 show the expected distribution as predicted by the theoretical analysis, that is, acceptance of both conditions in Catalan, which confirms the optionality of the SN *no*. In Spanish, there is a strong acceptance of the NCI-SN condition and rejection of the NCI+SN, confirming the lack of optionality of sentential negation with preverbal NCIs. Recall that all DOM targets in Spanish only require the accusative *a* marker and therefore, what is reported as -DOM is when the *a* is missing (ungrammatical in Spanish, yet the only grammatical in Catalan) and +DOM is when the *a* is present (grammatical in Spanish and ungrammatical in Catalan). The results for the -DOM condition in Catalan indicate target-like performance for both groups, however, both groups have high acceptance of the

**TABLE 2 |** Raw count of acceptance by condition for the two properties and two groups<sup>a</sup>.

	Catalan		Spanish	
NCI CONDITIONS				
	NCI+SN	NCI-SN	NCI+SN	NCI-SN
L1Cat-L2Sp ( <i>n</i> = 21)	127/168	141/168	36/168	147/168
L1Sp-L2Cat ( <i>n</i> = 23)	147/184	163/184	35/184	160/184
DOM CONDITIONS				
	−DOM	+DOM	−DOM	+DOM
L1Cat-L2Sp ( <i>n</i> = 21)	112/168	106/168	72/168	162/168
L1Sp-L2Cat ( <i>n</i> = 23)	129/184	114/184	51/164	168/184

<sup>a</sup>The total percentage of “I do not know” responses is 0.014%.

<sup>10</sup>We present the results collapsed in a binary coding for ease of exposition and to make a clearer distinction between grammatical and ungrammatical. A similar analysis was conducted on the 1-to-6 scale data and the overall picture was the same.



+DOM condition in Catalan (ungrammatical). When the participants are tested in Spanish, they each show target-like acceptance of the +DOM sentences, but, in the ungrammatical condition (-DOM), they also show a slight over-acceptance.

### Statistical Analysis

To further investigate the findings, we conducted linear mixed effects logistic regression analyses of the responses in the R environment (R Core Team, 2016), by using the lme4 package (Bates et al., 2015). Generalized mixed effects models were fit to the binomial response data. The data for the two properties under investigation were analyzed separately in each language, thus, we used separate models. Each model included fixed effects of condition (Model1: NCI+SN, NCI+V; Model2: -DOM, +DOM), group (L1Cat-L2Sp, L1Sp-L2Cat,) and their interaction. Fixed effects were sum-coded as -0.5/0.5 and each model included by-participant and by-item random intercepts and slopes for the repeated measures variables. In the case of significant interactions, planned comparisons investigated effects of group within the same condition using the multcomp package (Hothorn et al., 2008). The summaries of the omnibus models are presented in **Tables 3, 4**.

For the NCI data the effect of condition was significant for the Spanish data only, in the absence of any significant interactions. This confirms that both groups allowed both conditions in Catalan and that both groups significantly preferred the NCI+V condition in Spanish. For the DOM data, there was a significant main effect of condition in both Catalan and Spanish, with a preference for the grammatical condition in each language.

The results show that both groups have target-like grammars in both Catalan and Spanish with respect to the NCIs. They all allow for optionality in the co-occurrence of the NCI (ningú) and Sentential Negation (no) as expected and they do not allow this optionality in Spanish. With regards to the DOM conditions,

both groups prefer the grammatical condition in each language +DOM in Spanish and -DOM in Catalan, but both groups also show an unexpected over-acceptance of ungrammatical conditions in both languages.

### Self-Paced Reading Tasks

Comprehension accuracy was calculated to ensure that participants were reading the sentences and paying attention to the task. The mean accuracy for the L1Cat-L2Sp group is 93.04% in Spanish and 95.61% in Catalan. The rates of comprehension accuracy for the L1Sp-L2Cat were 92.30% in Spanish and 94.31% in Catalan. This indicates that the participants paid attention to the task.

The analysis focuses on the three regions following the Critical Region to check for any slowing down effects (i.e., spill-over effects). This is done due to the fact that for two of the four conditions, the Critical Region was an empty region (absence of Sentential Negation or absence of the accusative marker). The reaction times (RTs) for each condition and each language were analyzed separately (NCI+SN, NCI-SN, +DOM, -DOM) using linear mixed effects models, using the same coding scheme as for the offline data. We used raw RT as opposed to residual because the critical comparisons are the same across conditions rendering residualization not necessary. The regions of interest were of the same length, both groups are equally bilingual (scoring at ceiling in proficiency in both languages), each bilingual group is highly literate in both languages and most crucially, there is purposefully no monolingual control comparison from which we might expect a general difference in reaction time. **Figure 1** shows the mean RTs (ms) for the three regions of interest and each group in the NCI conditions when the groups are tested in Catalan.

The three models revealed no significant main effects or interactions (see **Table 5**), which indicates that both the L1Cat-L2Sp and the L1Sp-L2Cat groups allow optionality with respect to the co-occurrence of pre-verbal NCIs and Sentential negation in Catalan.

When they are tested in Spanish in these same conditions, the picture that emerges is different (see **Figure 2**).

As seen in **Table 6**, the only significant main effect was the one on condition in R2, showing that both groups are significantly slower in the second region of interest of the NCI+SN (ungrammatical in Spanish) than in the NCI+V (grammatical). The results show that both groups are sensitive to the morphosyntactic violation of pre-verbal NCIs co-occurring with Sentential Negation in Spanish.

Turning to the DOM conditions, **Figure 3** illustrates the Catalan data.

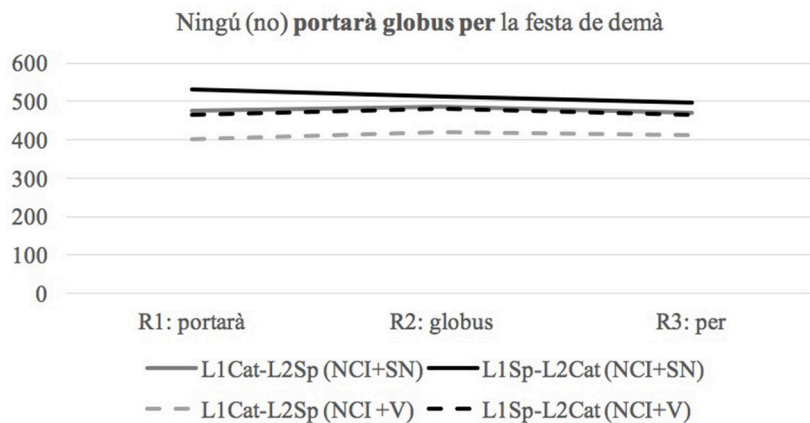
The statistical results in **Table 7** show that there is a significant interaction of Group\*Condition in the third region of interest. The results indicate that the L1Cat-L2Sp group does not show sensitivity to the morphosyntactic violation of the +DOM condition and the L1Sp-L2Cat group shows sensitivity to the -DOM condition, being significantly slower in the first ( $p = 0.025$ ) and third region ( $p < 0.001$ ). This shows that the L1Cat-L2Sp group has optionality in their grammars because they allow sentences with the accusative marker and without it in Catalan and that the L1Sp-L2Cat disallows the absence of the

**TABLE 3 |** Generalized mixed effects models for the NCI property in the two different datasets (RL: L1Cat-L2SP, NCI+SN).

	Catalan data		Spanish data	
	Estimate (SE)	<i>p</i>	Estimate (SE)	<i>p</i>
(Intercept)	1.094 (0.28)	< 0.001	-1.338 (0.44)	< 0.001
Group	0.251 (0.25)	0.316	-0.181 (0.44)	0.488
Condition	0.726 (0.43)	0.092	3.278 (0.55)	< 0.001
Group: Condition	0.334 (0.42)	0.418	0.413 (0.46)	0.322

**TABLE 4 |** Generalized mixed effects models for the DOM property in the two different datasets (RL: L1Cat-L2SP, -DOM).

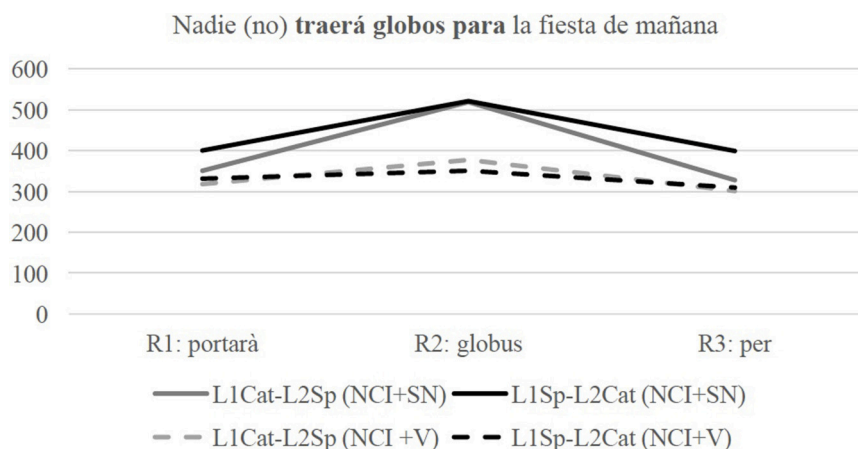
	Catalan data		Spanish data	
	Estimate (SE)	<i>p</i>	Estimate (SE)	<i>p</i>
(Intercept)	1.815 (0.28)	< 0.001	-1.092 (0.44)	< 0.001
Group	-0.114 (0.34)	0.738	-0.216 (0.38)	0.039
Condition	-1.627 (0.34)	< 0.001	2.877 (0.54)	< 0.001
Group: Condition	0.595 (0.39)	0.136	0.022 (0.39)	0.954



**FIGURE 1** | Line graph of Reaction Times (RTs) in milliseconds (ms) in the regions of interest in the NCI conditions when tested in Catalan.

**TABLE 5** | Linear models for the NCI property Catalan (RL: L1Cat-L2SP, NCI+SN).

	R1		R2		R3	
	Estimate (SE) t-value	P	Estimate (SE) t-value	p	Estimate (SE) t-value	p
Intercept	452.1 (32.7) 13.809	<0.001	463.9 (32.1) 14.459	<0.001	394.5 (26.6) 14.822	<0.001
Condition	−81.6 (50.4) 1.619	0.105	−45.55 (40.7) −1.118	0.263	−85.6 (45.5) −1.882	0.069
Group	77.3 (46.7) 1.654	0.09	66.9 (52.06) 1.654	0.198	66.8 (38.1) 1.753	0.079
Condition*Group	−60.6 (42.1) 1.439	0.149	−07.6 (31.6) −1.439	0.809	−82.3 (52.5) −1.566	0.117



**FIGURE 2** | Line graph of Reaction Times (RT) in milliseconds (ms) in the regions of interest in the NCI conditions when tested in Spanish.

accusative marker, potentially showing influence from Spanish onto Catalan.

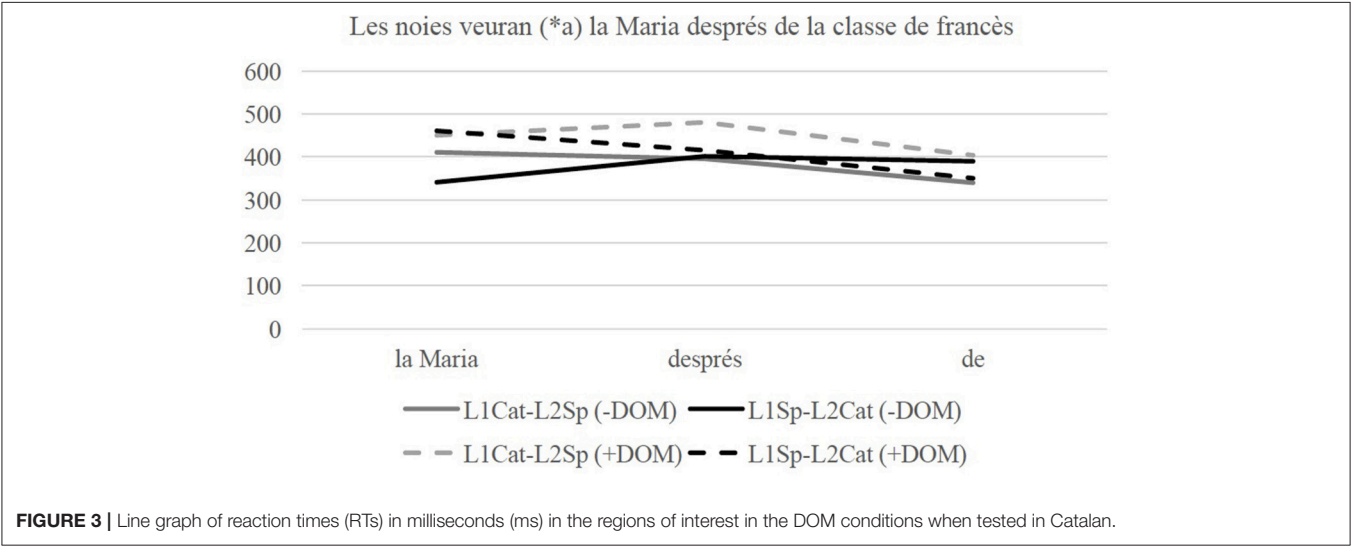
The following **Figure 4** shows the Spanish Data in the DOM conditions.

The statistical models in **Table 8** show a significant interaction of Group\*Condition in Region 1, reflecting that the L1Cat-L2Sp

group is significantly slower in the +DOM condition ( $p < 0.001$ ) and the L1Sp-L2Cat group is significantly slower in the −DOM condition. In the third region, there is also significant interaction of Group\*Condition, the L1Sp-L2Cat group is significantly slower in the −DOM condition. Overall, the results show that the group of L1Sp-L2Cat group have target-like grammar and

**TABLE 6 |** Linear models for the NCI property in Spanish (RL: L1Cat-L2SP, NCI+SN).

	R1		R2		R3	
	Estimate (SE) t-value	p	Estimate (SE) t-value	P	Estimate (SE) t-value	P
Intercept	399.2 (38.6) 10.329	<0.001	452.5 (17.3) 26.164	<0.001	335.3 (18.5) 18.092	<0.001
Condition	1.45 (60.5) 0.025	0.979	−190.6 (19.1) −9.995	<0.001	−46.4 (28.2) −1.693	0.092
Group	113.69 (77.3) 1.471	0.141	−26.1 (34.3) −0.759	0.447	51.8 (34.6) 1.496	0.134
Condition*Group	50.81 (120.9) 0.420	0.420	−44.1 (36.9) −1.192	0.233	−58.9 (49.79) −1.192	0.236



**TABLE 7 |** Linear models for the DOM property in Catalan (RL: L1Cat-L2SP, −DOM).

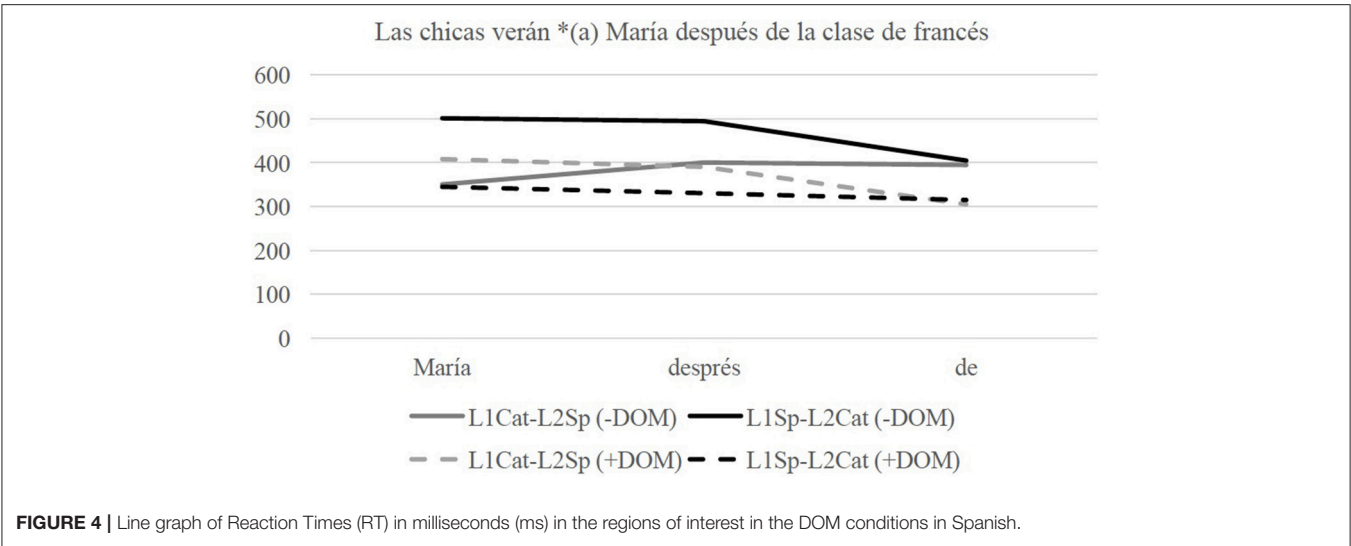
	R1		R2		R3	
	Estimate (SE) t-value	p	Estimate (SE) t-value	p	Estimate (SE) t-value	p
Intercept	454.2 (29.7) 15.265	<0.001	455.1 (27.7) 16.411	<0.001	371.4 (22.7) 22.674	<0.001
Condition	−19.9 (48.3) −0.415	0.678	−15.5 (49.9) −0.310	0.756	15.3 (26.33) 0.581	0.561
Group	65.9 (41.4) 1.583	0.113	45.4 (29.8) 1.515	0.129	1.44 (26.36) 0.054	0.956
Condition*Group	−98.4 (44.1) −2.232	0.025	−196.9 (35.8) −5.496	0.388	−121.18 (35.5) −3.410	<0.001

that the L1Cat-L2Sp group show sensitivity to the expected grammatical condition, thus, their grammar shows influence from Catalan with respect to this phenomenon in Spanish.

## DISCUSSION AND CONCLUSIONS

In this section, we bring the results together in summary. As there is a significant amount of data to be considered, we begin with a brief overview of the most interesting results. Starting

with the NCI conditions, as can be seen in **Table 9**, irrespective of modality (offline vs. online) and the language of testing, each group's performances are consistent with having distinct representations for both languages that conform to what is formally described of Spanish and Catalan. As a result we can safely say that order of acquisition and/or relative dominance in one or the other language brings nothing to bear for this domain of grammar, at least for these sets of bilinguals, a point to which we return below.



**TABLE 8 |** Generalized linear models for the DOM property in Spanish (RL: L1Cat-L2SP, -DOM).

	R1		R2		R3	
	Estimate (SE) t-value	p	Estimate (SE) t-value	p	Estimate (SE) t = value	p
Intercept	406.7 (43.8) 9.273	<0.001	365.9 (21.7) 16.846	<0.001	355.3 (18.8) 18.805	<0.001
Condition	15.3 (26.3) −0.619	0.535	−101.3 (30.7) −3.297	<0.001	−63.4 (31.5) −2.010	0.027
Group	44.1 (87.7) 0.502	0.615	75.3 (42.4) 1.774	0.076	4.89 (35.1) 0.139	0.889
Condition*Group	−232.3 (35.5) −1.365	<0.001	−84.5 (58.5) −1.443	0.148	27.5 (56.4) 0.487	0.625

Turning to the DOM conditions, the picture is less clear. We have some within group mismatches in performance across modalities as well as inter-group across language and modality-group, as can be appreciated visually in **Table 10** below. Our focus is definitively not on any comparisons to monolinguals, but rather on a fairer bilingual-to-bilingual group comparison (e.g., Ortega, 2010, 2013; Rothman and Iverson, 2010; Hopp and Schmid, 2013) where L1 and L2 status is switched in a mirror-image way and proficiency is held constantly high in. That said, we do highlight below where group diverges from expected monolingual norms, as described in the literature, with some insights as to why this might be. Attempting to compare the bilinguals to monolingual control groups would have been difficult, in part since it would be virtually impossible to find a Catalan monolingual control group and thus it would have been unbalance if we were to offer only a Spanish one. At first glance, however, it is useful to highlight, as we predicted could occur, that CLI can be conditioned by the domain of grammar itself, a point to which we will return in greater detail below.

Looking at the quadrant on the top-left side of the table shaded in green, that is when L1-Sp-L2-Cat bilinguals are tested in their L1, Spanish, we see that for the -DOM conditions—where the accusative marker *a* is not present although it is grammatically

**TABLE 9 |** Summary of the results for the NCI conditions, where (✓) refers to expected performance based and (✗) does not.

	L1Sp-L2Cat		L1Cat-L2Sp	
	NCI+SN (GJT)	NCI-SN (GJT)	NCI+SN (GJT)	NCI-SN (GJT)
Spanish	✓	✓	✓	✓
	NCI+SN (SPR)	NCI-SN (SPR)	NCI+SN (SPR)	NCI-SN (SPR)
Catalan	✓	✓	✓	✓
	NCI+SN (SPR)	NCI-SN (SPR)	NCI+SN (SPR)	NCI-SN (SPR)

obligatory—the GJT revealed influence from Catalan, their L2. This is not terribly surprising in light of previous literature that has shown DOM to be highly vulnerable in bilingual contexts (e.g., Guijarro-Fuentes and Marinis, 2009; Montrul et al., 2015) even for the context we used—purposefully because dialectal variation that can otherwise obtain for DOM in other contexts does not apply. However, it is not clear at what level this Catalan influence rests—e.g., if such reflects a representational difference in their mental grammars—precisely because in the SPR task the



**TABLE 10 |** Summary of the results for the DOM conditions, where (✓) refers to expected performance based and (✗) does not.

	L1Sp-L2Cat		L1Cat-L2Sp	
Spanish	+DOM (GJT)	–DOM (GJT)	+DOM (GJT)	–DOM (GJT)
	✓	✗	✓	✗
	+DOM (SPR)	–DOM (SPR)	+DOM (SPR)	–DOM (SPR)
	✓	✓	✓	✗
Catalan	+DOM (GJT)	–DOM (GJT)	+DOM (GJT)	–DOM (GJT)
	✗	✓	✗	✓
	+DOM (SPR)	–DOM (SPR)	+DOM (SPR)	–DOM (SPR)
	✗	✗	✗	✓

same participants do show a clear sensitivity to the very same ungrammatical condition. If it were truly the case that these speakers' grammars did not have the functional architecture of Spanish DOM in their grammar, we would expect that they would be equally insensitive to DOM grammaticality issues in both modalities. The fact that the processing measure shows sensitivity that is potentially obscured in the offline behavioral measures alone might be because the processing measures are more likely to tap into implicit knowledge (e.g., Jegerski, 2014; Keating and Jegerski, 2015). Therefore, we would not conclude based on a coupling of the two modalities that these L1-Sp-L2-Cat bilinguals have non-monolingual-like representations for DOM, but rather that the offline task shows a more methodological performance based difficulty. This same pattern, where processing measures indicate better competence than offline behavioral measures, has been shown recently for other types of Spanish bilinguals, namely more traditional HSs in North America (e.g., Villegas, 2014; Jegerski et al., 2016).

Shifting to the bottom-left quadrant of the table shaded in blue, that is when the L1-Sp-L2-Cat bilinguals are tested in Catalan, they show over-acceptance of sentences with +DOM (ungrammatical in Catalan) in the GJT and they do not show sensitivity to the morphosyntactic violation in this condition in the SPR either. Because there is performance conformity across modalities, we take this as especially strong evidence that the underlying reason for both performances is one and the same, that is, representational in nature. The performance seems to suggest that Spanish is influencing their Catalan. In turn, their performance in Catalan as summarized in **Table 10** is further evidence for what we argued in relation to the representation of this domain in their Spanish grammar. Recall that they appeared to have some issues marking –DOM as ungrammatical despite having no issues accepting +DOM as grammatical and being sensitive to the –DOM violation in RT. We concluded that the processing measure reflected their competence more accurately. Their performance on the Catalan condition thus seems to strengthen this claim precisely because one could only reasonably expect (or explain) evidence of Spanish DOM transfer in Catalan if indeed they had an intact DOM representation from their other grammar. There is also a modality asymmetry in their Catalan performance for the same domain, that is –DOM, however, this seems to be the mirror image of their performance in Spanish. In Catalan, they perform just fine in the –DOM condition, which

entails accepting as grammatical sentences that do not have an overt *a* case marker, in the offline measure only. With the same condition in the online measure, they show a sensitivity (they slow down) where they should not, suggesting that they are sensitive to a grammatical violation that should not obtain in Catalan but does in Spanish. We would like to suggest that the offline measure potentially reflects a “yes” bias, they simply did not reject something provided to them and that the online measure reflects more their grammatical representation, which we take to be influenced from Spanish. To the extent that this is on the right track, it again provides further evidence for intact DOM representations in Spanish.

These results, related back to our research question that probes the relationship that language use and exposure exercises on linguistic competence/performance in both languages of early child bilinguals, suggest that language use and exposure play a role in determining the directionality of cross-linguistic influence.<sup>11</sup> Recall that this set of participant was categorized as having high use and exposure to Spanish even though they live in a Catalan-dominant area. We conclude that contrary to other typical cases of Spanish Heritage Speaker bilingualism, the access to high quality and quantity of input to the minority language of the immediate context (i.e., Spanish)—by means of language use and exposure on top of education—is a key factor to preventing cross-linguistic interference from the majority language of the immediate context (i.e., Catalan).

Turning to the L1-Cat-L2-Sp bilinguals, we focus our attention to the quadrant on the top-right of the table shaded in orange. Particularly notable is the fact that they do not judge the –DOM conditions in Spanish as categorically ungrammatical (GJT), nor do they show appropriate sensitivity to the ungrammaticality in this condition. However, in the +DOM conditions, they show target-like performance in the GJT and SPR tasks. Since it is the case that these bilinguals do not reliably reject nor show sensitivity in RT to sentences in Spanish without the accusative *a* marking when the object is [+animate, +specific], the canonical condition under which DOM is required, yet have no issues accepting sentences that have it in the same context, we might conclude that they indeed have a representation for DOM in their mental Spanish grammars, but, unlike the other group and other sets of Spanish natives described in the literature, DOM seems optional as opposed to obligatory. Such a conclusion might be strengthened by the latent patterns in their performance. That is, in both the –DOM and +DOM they are consistent in their performances across offline and online modalities.

Turning to the final quadrant in the bottom right shaded in yellow, that is, when the L1-Cat-L2-Sp participants are tested in

<sup>11</sup>The two languages under investigation are two closely related systems and thus, this might have had an effect on triggering cross-linguistics effects. However, cross-linguistic influence in the DOM we have investigated has also been reported when Spanish is in context with other less related language, such as English, in context of Spanish as a Heritage Language in the US (e.g., Montrul, 2004; Montrul and Bowles, 2009; Montrul and Sánchez Walker, 2013) or Spanish as a non-native language (e.g., Farley and McCollam, 2004; Bowles and Montrul, 2008) in North America (e.g., Montrul, 2004; Montrul and Bowles, 2009; Montrul and Sánchez Walker, 2013).

their native Catalan, we see that although they prefer sentences without DOM (grammatical in Catalan) by rating them as more acceptable than sentences with DOM (ungrammatical in Catalan), they do accept +DOM sentences at a non-trivial rate. In the online data, these speakers show no sensitivity in –DOM conditions, as expected, however, they do not show sensitivity to the grammatical violation of +DOM conditions in Catalan. Taken together, this also suggests that their grammars allow for optionality with respect to DOM, which goes in line with Chondrogianni (2018) claim that DOM in Catalan is starting to appear in varieties of Catalan which traditionally do not allow for it. Optionality in their Spanish grammar, thus, can be explained by influence from Catalan on their Spanish precisely because their Catalan shows the same degree of optionality. As it relates to the question of language dominance (LD), again we see that LD affects cross-linguistic influence in these highly proficient bilinguals. At first glance, because there is optionality that would not be expected *per se* of a monolingual native Catalan grammar (to the extent that there are any) it was not clear that LD, in this case Catalan influence, was unambiguously demonstrated or at least as clearly as it was for the Spanish dominant group. However, since we have shown that the optionality in Spanish is reflected also in the Catalan of these same speakers it seems reasonable to understand the optionality in Spanish as influence of Catalan as represented in these bilinguals.<sup>12</sup> Thus, we have evidence of LD affecting cross-linguistic influence in both groups.

It is interesting to ponder why out of the two domains of grammar tested, both of which differ across the languages, only one shows cross-linguistic influence, albeit patterning differently, in both target groups. It is possible that the issues with DOM are idiosyncratic to DOM itself. Recall that DOM seems to be challenging in all instances of heritage speaker bilingual acquisition (see e.g., Montrul et al., 2015). Moreover, we should keep in mind that the accusative case marker itself is phonologically reduced and potentially not overly salient. Furthermore, DOM reflects a large degree of variation across

Spanish dialects and even individual speakers. Because our bilinguals, however, are all exposed to Peninsular Spanish where DOM is consistent in the core context we isolated (López, 2012) and given that the [+animate, +specific] is not subject anyway to much variation dialectally or individually, we attempted to control for the general variation within this grammatical domain, which was chosen precisely because it had been shown to be problematic for more typical HSs. Keeping in mind our research questions then and under the hypothesis that less variation would obtain in our context of societal bilingualism as compared to more traditional HS situations, examining a domain such as DOM, as compared to other properties, could then go a long way to inform us about what is vulnerable in bilingualism even when many variables that likely affect HS performance are more favorably proportioned. And so, why all of these factors might contribute to why DOM is a vulnerable property for bilingual variation in general, they do not seem to be overcome as they are for monolinguals even in an environment where all opportunity has been given for our bilinguals to perform like monolinguals. This should of course not be surprising and certainly bespeaks nothing evaluative about our bilinguals herein, why would they or how could they perform exactly like monolinguals, if only because they are simply not monolinguals? However, given the differences across the two groups that grow up under similarly favorable environmental conditions, there does seem to be some evidence to suggest that order of acquisition/language dominance matters for the outcomes of development in this domain. And so, relating more directly all that we have seen across the two domains of grammar to our two research questions, it seems that LD matters for some domains of grammar more than others, even when bilinguals are more or less balanced as related to overall proficiency in the languages and when this is maximally supported by a bilingual environment. If the same pattern holds for future studies of a similar nature, then looking at the adult outcomes of such groups as we have done here might couple together with more traditional HS populations to inform linguistic theory more generally. As Polinsky (2016, 2018) has nicely argued and supported with data recently, certain domains of grammar are invulnerable to bilingual effects even in the minority language of HSs who are severely imbalanced in dominance whereas others are highly sensitive to bilingual effects. Our data then support her general claim (see Tsimpli, 2014 for similar arguments), showing that some properties of grammar are still vulnerable to bilingual effects while others are not even in the opposite case, that is, when there is extremely high proficiency in both languages and the day-to-day environment of the bilingual promotes both languages. Together, such data can tell us what is more and less core related to language in general.

As promised above, it is worth coming back to the case of NCIs and ponder why there is no CLI noted at all, that is, conditioned or not by order of acquisition/dominance, different from the case of DOM. The case of NCIs is interesting by comparison to DOM, since only the former relates to optionality in the “larger” grammar. Catalan permits the Spanish sole, obligatory spell-out [the use of the NCI without Sentential Negation (SN)] but optionally allows for double negation spell-out without the canceling of semantic negation (as would be

<sup>12</sup>One reviewer queried whether or not subtle differences, as we have uncovered herein, in bilingual grammars would serve as a potential catalyst for changing the representational structure of monolingual grammars. It is outside of the scope of this paper to make such claims, not the least because it is difficult to find monolinguals of Catalan in particular to test what we would claim. That said, a general discussion on the matter is perhaps warranted. As monolinguals are in contact with bilinguals, especially in situations like Catalonia where bilingualism is the default state—rural enclaves of monolingualism would likely have significant contact with bilinguals, in person or via media. It would, thus, make sense that bilingual innovations could result in changes to monolinguals via various paths. We will highlight one herein. In light of L1 attrition research (see Schmid and Köpcke, 2017, for review), we know that native grammars can change over time. We also know from Iverson (2012) and Iverson and Miller (2017) that all domains of grammar—even narrow syntax—can be affected by shifting input over thresholds for L1 change over time. And so, contact with bilinguals over the lifespan can induce innovations—if the threshold is tipped. Changes in production as a result in monolingual grammars will likely affect how the next generation sets the grammatical system, as argued for monolingual L1 acquisition (Lightfoot, 1999) and heritage speakers (Pascual y Cabo and Rothman, 2012; Bayram et al., 2018), albeit via somewhat distinct provenance. Our results would be compatible with the argumentation in Perpiñán (2018) that specific context of Catalan-Spanish bilingualism is leading to language change and to the creation of a new variety of Catalan that allows for the optionality seen in our participants.

the case in Spanish if an NCI co-occurred with SN). And so, there is no direct competition of an obligatory nature between the two grammars, as is the case with DOM where an obligatory use of DOM constitutes an ungrammatical extension of DOM in Catalan. Therefore, it could be the case that this tension “optionality” vs. “obligatoriness” plays a further conditioning role for CLI. In a sense, the grammars might be less likely to affect one another when what is at stake is not a contradiction in the obligatory construction of a grammatical structure. The subtleties involved, in other words, are actually not so subtle. The case of NCI might stand out across the two languages as more salient precisely because Catalan optionality coincides with a very specific domain of distribution in which it reflects an interpretation that is unavailable in Spanish.

As a closing point of discussion, it is worth considering whether or not our speakers are indeed HSs of a specific subtype or if it would indeed be best to not apply that label to them. In an effort to not open up Pandora’s box on this potential issue, we were neutral in distinguishing traditional HSs from our bilinguals herein and mainly because it hardly matters for our immediate points. We could be neutral because there is no denying the fact that our bilinguals are quite different in non-trivial ways from Spanish HSs studied in North America. But those differences alone do not necessarily mean that they are both not HSs, yet of distinct types (see Putnam et al., 2018 for similar argumentation). Although more traditional HSs do not remain dominant in their HL because their environments essentially preclude this and it is seemingly a given that HSs will show, on a gradient, differences from expected monolingual baselines (but see Kupisch and Rothman, 2016), a lack of difference in these regards should not be used as a criterion to disqualify someone as a HS. Doing so would only make sense under a deficit model of HS bilingualism whereby the label HS has somehow become synonymous with deficiency par excellence. With many others (e.g., Putnam and Sánchez, 2013; Kupisch and Rothman, 2016; Bayram et al., 2017; Putnam et al., 2018), we definitively reject such a view. Allowing for the present bilingual groups to be considered as a specific subtype of HSs, precisely because they meet all the neutral inclusion criteria of several non-deficit approaches definitions widely adopted in the literature, for example, Rothman (2009). And so, evidence from highly balanced HSs, if the label is appropriate to apply to our L1-Sp-L2-Cat group, could go a long way at counterbalancing the HS as an incomplete acquirer viewpoint. Our L1-Sp-L2-Cat participants grew up in a household where both parents had

moved to rural Catalonia and are not native speakers of Catalan, Spanish is their exclusive L1 and the only language spoken in their homes when they were young children and continues to be the family language. Crucially, the majority language of the immediate environment they grow up in is not their home language, but rather (for them) an L2 (Catalan), which they became significantly immersed in only upon going to school. This means that Spanish is their native L1, unlike the L1-Cat-L2-Sp group for whom Spanish is clearly an L2. It is also true that in this environment successful bilingualism and support for such is omnipresent and, thus, the possibility to maintain and further develop Spanish is different than other typical cases of HSs. Spanish has a higher prestige and is more accessible than it is in the USA, however, in this specific part of Catalonia there is no question that Spanish is not the majority language of the society (see Illamola, 2015). The increased opportunity to conserve dominance in Spanish does not disqualify our HSs from being HSs, it merely naturally creates an environment in which we can observe the relative weight of key variables that are different from Spanish HS situations in other environments and could not otherwise be teased apart. And so, why should our population not reflect a sub-type of HS? We leave this discussion for future work that takes advantage more and more of what comparisons of traditional HSs and bilinguals like ours can show when the minority language, in this case Spanish, is able to be held constant.

## AUTHOR CONTRIBUTIONS

EP-M is the main author. JR is the second main author and lab director. IC worked on the statistical analysis. FB, DM and ST worked on the conceptualization, design and implementation of the study.

## FUNDING

We thank the following funding bodies: the Language, Development and Aging Division at the University of Reading (RF: G16-142), Advancing the European Multilingual Experience (AthEME), funded by the European Seventh Framework Programme for research, technological development and demonstration under grant 613465; the following grants: FFI2014-52015-P and FFI2017-82547-P (Spanish MINECO); and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Individual Fellowship grant agreement No. [799652].

## REFERENCES

- Bates, D., Maechler, M., Bokler, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bayram, F., Pascual y Cabo, D., and Rothman, J. (2018). “cross-generational attrition contributions to heritage speaker competence,” in *The Oxford Handbook of Attrition*, eds B. Köpke and M. S. Schmid (Oxford: Oxford University Press).
- Bayram, F., Rothman, J., Iverson, M., Kupisch, T., Miller, D., Puig-Mayenco, E., et al. (2017). Differences in use without deficiencies in competence: passives in the Turkish and German of Turkish heritage speakers in Germany. *Int. J. Biling. Educ. Biling.* 32, 1–27. doi: 10.1080/13670050.2017.1324403
- Bel, A. (1996). Early negation in Catalan and Spanish. *Catal. Work. Pap. Linguist.* 5, 5–28.
- Bel, A. (2001). *Teoria Lingüística i Adquisició del Llenguatge: Anàlisi Comparada dels Trets Morfològics en Català i Castellà*. Barcelona: Institut d’Estudis Catalans.



- Bel, A. (2003). The syntax of subjects in the acquisition of Spanish and Catalan. *Probus* 15, 1–26. doi: 10.1515/prbs.2003.003
- 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
- Bialystok, E. (2007). Cognitive effects of bilingualism: how linguistic experience leads to cognitive change. *Int. J. Biling. Educ. Biling.* 10, 210–233. doi: 10.2167/beb441.0
- Bosch, L., and Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy* 1, 29–49. doi: 10.1207/S15327078IN0201\_3
- Bowles, M., and Montrul, S. (2008). “The role of explicit instruction in the L2 acquisition of the a-personal,” in *Selected Papers from the 8th Hispanic Linguistics Symposium*, eds J. Almazán, J. Bruhn de Garavito, and E. Valenzuela (Somerville, MA: Cascadia Press), 25–35.
- Bruhn de Garavito, J., and Valenzuela, E. (2008). Eventive and stative passives in Spanish L2 acquisition: a matter of aspect. *Biling. Lingu. Cognit.* 11, 323–336. doi: 10.1017/S1366728908003556
- Chondrogianni, V. (2018). “Child L2 acquisition,” in *Bilingual Cognition and Language: The State of the Science Across Its Subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins), 103–126.
- Déprez, V., Tubau, S., Cheylus, A., and Espinal, M. T. (2015). Double negation in a negative concord language: an experimental investigation. *Lingua* 163, 75–107. doi: 10.1016/j.lingua.2015.05.012
- Escandell-Vidal, V. (2009). Differential object marking and topicality: the case of Balearic Catalan. *Stud. Lingu.* 33, 832–885. doi: 10.1075/sl.33.4.02esc
- Espinal, M. T. (2000). On the semantic status of n-words in Catalan and Spanish. *Lingua* 110, 557–580. doi: 10.1016/S0024-3841(00)00002-4
- Espinal, M. T., and Tubau, S. (2016). “Meaning of words, meaning of sentences. Building the meaning of n-words,” in *Manual of Grammatical Interfaces in Romance* eds S. Fischer and C. Gabriel (Berlin: De Gruyter), 187–212.
- Espinal, M. T., Tubau, S., Borrás-Comes, J., and Prieto, P. (2016). “Double negation in Catalan and Spanish. Interaction between syntax and prosody,” in *Negation and Polarity: Experimental Perspectives*, eds P. Larrivé and C. Lee (Zürich: Springer International Publishing), 145–175.
- Farkas, D. (1995). “Specificity and scop,” in *Actes du Premier Colloque Langues & Grammaire*, eds L. Nash and G. Tsoulas (Paris: Département des sciences du langage, Université de Paris), 119–137.
- Farley, A., and McCollam, K. (2004). Learner readiness and L2 production in Spanish: processability theory on trial. *Estud. Lingüíst. Aplic.* 22, 47–69. Available online at: <https://dialnet.unirioja.es/servlet/articulo?codigo=6395879>
- GIEC (2016). *Gramàtica de la Llengua Catalana*. Barcelona: Institut d'Estudis Catalans.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain Lingu.* 36, 3–15. doi: 10.1016/0093-934X(89)90048-5
- Guijarro-Fuentes, P. (2012). The acquisition of interpretable features in L2 Spanish: personal a. *Biling. Lingu. Cognit.* 15, 701–720. doi: 10.1017/S1366728912000144
- Guijarro-Fuentes, P., and Marinis, T. (2009). “The acquisition of personal preposition a by Catalan-Soanish and English-Spanish bilinguals,” in *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, eds J. Collentine, B. Lafford, and F. Marcos Marín (Somerville, MA), 81–92.
- Haegeman, L., and Zanuttini, R. (1991). Negative heads and the Neg criterion. *Linguist. Rev.* 8, 233–252. doi: 10.1515/tlir.1991.8.2-4.233
- Haznedar, B. (2013). Child second language acquisition from a generative perspective. *Linguist. Approach. Biling.* 3, 26–47. doi: 10.1075/lab.3.1.02haz
- Hopp, H., and Schmid, M. (2013). Perceived foreign accent in L1 attrition and L2 acquisition: the impact of age of acquisition and bilingualism. *Appl. Psycholinguist.* 34, 361–394. doi: 10.1017/S0142716411000737
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biomet. J.* 50, 346–363. doi: 10.1002/bimj.200810425
- Idescat (2013). *Enquesta d'usos Lingüístics de la Població 2013*. Barcelona, Spain.
- Illamola, C. (2015). *Contacto de Lenguas y la Expresión de la Posterioridad Temporal en Español de Cataluña*. Barcelona: Universitat de Barcelona.
- Iverson, M. (2012). *Advanced Language Attrition of Spanish in Contact with Brazilian Portuguese*. University of Iowa.
- Iverson, M., and Miller, D. (2017). Language attrition and maintenance. *Linguist. Approach. Biling.* 7, 704–708. doi: 10.1075/lab.00009.ive
- Jegerski, J. (2014). “Self-paced reading,” in *Research Methods in Second Language Psycholinguistics*, eds J. Jegerski and B. VanPatten (New York, NY: Routledge), 20–49.
- Jegerski, J., Keating, G., and VanPatten, B. (2016). On-line relative clause attachment strategy in heritage speakers of Spanish. *Int. J. Biling.* 20, 254–268. doi: 10.1177/1367006914552288
- Karayayla, T. (2018). *Turkish as an Immigrant and a Heritage Language in the UK: Effects of Exposure and Age of Onset of Bilingualism on Grammatical and Lexical Development of the First Language*. Colchester: University of Essex.
- Karayayla, T., and Schmid, M. (2017). “The role input for language development in Heritage Speakers - quality vs. quantity,” in *Paper Presented at EuroSLA 27*, (University of Reading, United Kingdom).
- Keating, G., and Jegerski, J. (2015). Experimental designs in sentence processing research: a methodological review and user's guide. *Stud. Second Lang. Acquis.* 37, 281–307. doi: 10.1017/S0272263114000187
- Kupisch, T., and Rothman, J. (2016). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Biling.* doi: 10.1177/1367006916654355. [Epub ahead of print].
- Laka, I. (1990). *Negation in Syntax: On the Nature of Functional Categories and Projections*. Cambridge, MA: Massachusetts Institute of Technology.
- Leonetti, M. (2004). Specificity and differential object marking in Spanish. *Catal. J. Linguist.* 3, 75–114. doi: 10.5565/rev/catjl.106
- Lightfoot, D. (1999). *The Development of Language: Acquisition, Change, and Evolution*. Oxford: Wiley-Blackwell.
- López, L. (2012). *Indefinite Objects. Scrambling, Choice Functions and Differential Marking*. Cambridge, MA: MIT Press.
- Marian, V., Blumfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (Leap-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 950–967. doi: 10.1044/1092-4388(2007/067)
- Meisel, J. (2011). *First and Second Language Acquisition*. Cambridge: Cambridge University Press.
- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: a case of morpho-syntactic convergence. *Biling. Lingu. Cogn.* 7, 125–142. doi: 10.1017/S1366728904001464
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: John Benjamins.
- Montrul, S. (2015). “Dominance and proficiency in early and late bilingualism,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35.
- Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press.
- Montrul, S., and Bowles, M. (2009). Back to basics: differential object marking under incomplete acquisition of Spanish heritage speakers. *Biling. Lingu. Cognit.* 12, 363–383. doi: 10.1017/S1366728909990071
- Montrul, S., and Sánchez Walker, N. (2013). Differential object marking in child and adult Spanish heritage speakers. *Langu. Acquisit.* 20, 109–132. doi: 10.1080/10489223.2013.766741
- Montrul, S., and Slabakova, R. (2003). Competence similarities between native and near-native speakers (an investigation of the preterit/imperfect contrast in Spanish. *Stud. Second Lang. Acquisit.* 25, 351–398. doi: 10.1017/S0272263103000159
- Montrul, S., Bhatt, R., and Girju, R. (2015). Differential object marking in Spanish, Hindi and Romanian as heritage languages. *Language* 91, 564–610. doi: 10.1353/lan.2015.0035
- Nicoladis, E. (2018). “Simultaneous Child Bilingualism,” in *Bilingual Cognition and Language: The State of the Science Across Its Subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins), 81–102.
- Ortega, L. (2010). “The bilingual turn in SLA,” in *Plenary Address at the American Association for Applied Linguistics Conference* (Atlanta, GA).
- Ortega, L. (2013). SLA for the 21st Century: disciplinary progress, transdisciplinary relevance, and the Bi/multilingual turn. *Lang. Learn.* 63, 1–24. doi: 10.1111/j.1467-9922.2012.00735.x



- Pascual y Cabo, D., and Rothman, J. (2012). The (il)logical problem of heritage speaker bilingualism and incomplete acquisition. *Appl. Linguist.* 33, 450–455. doi: 10.1093/applin/ams037
- Perpiñán, S. (2017). Catalan-Spanish bilingualism continuum: the expression of non-personal Catalan clitics in the adult grammar of early bilinguals. *Linguist. Approac. Biling.* 7, 477–513. doi: 10.1075/lab.15004.per
- Perpiñán, S. (2018). Ongoing language change, and crosslinguistic influence in direct object expression in Catalan-Spanish bilingualism. *Languages* 3, 1–14. doi: 10.3390/languages3020014
- Polinsky, M. (2016). “Cascading consequences of syntactic reorganization: Ellipsis in heritage languages,” in *Keynote Address at BUCLD 41* (Boston, MA).
- Polinsky, M. (2018). *Heritage Languages and their Speakers*. Cambridge: Cambridge University Press.
- Putnam, M., and Sánchez, L. (2013). What’s so incomplete about incomplete acquisition?: a prolegomenon to modeling heritage language grammars. *Linguist. Approac. Biling.* 3, 478–508. doi: 10.1075/lab.3.4.04put
- Putnam, M., Kupisch, T., and Pascual y Cabo, D. (2018). “Different situations, similar outcomes: heritage grammars across the lifespan,” in *Bilingual Cognition and Language: The state of the science across its subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins Publishing Company), 251–279.
- Quer, J. (1993). *The Syntactic Licensing of Negative Elements*. MA thesis, Universitat Autònoma de Barcelona.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna: Available online at: <http://www.R-project.org>.
- Rodríguez-Mondoñedo, M. (2007). *The syntax of Objects: Agree and Differential Object Marking*. Ph.D. dissertation. University of Connecticut.
- 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 Iverson, M. (2010). “Independent normative assessments for Bi/Multilingualism, where art thou?” in *Multilingual Norms* ed M. Cruz-Ferreira (Frankfurt: Peter Lang Publishing), 33–51.
- Rothman, J., and Slabakova, R. (2017). The generative approach to SLA and its place in modern second language studies. *Stud. Second Lang. Acquis.* 40, 471–442. doi: 10.1017/S0272263117000134
- 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
- Schmeißer, A., Hager, M., Arnaus Gil, L., Jansen, V., Geveler, J., Eichler, N. A., et al. (2015). “Related but different: the two concepts of language dominance and language proficiency,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 36–65.
- Schmid, M. S., and Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguist. Approac. Biling.* 7, 637–667. doi: 10.1075/lab.17058.sch
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: determinants and mechanisms. *Linguist. Approac. Biling.* 3, 3–25. doi: 10.1075/lab.3.1.01ser
- Silva-Corvalán, C., and Treffers-Daller, J. (2015). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press.
- Simonet, M. (2011). Production of Catalan-specific vowel contrast by Spanish-Catalan bilinguals. *Phonetica* 68, 88–110. doi: 10.1159/000328847
- Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *J. Phon.* 43, 26–37. doi: 10.1016/j.wocn.2014.01.004
- Slabakova, R., and Montrul, S. (2003). Genericity and aspect in L2 acquisition. *Lang. Acquis.* 11, 165–196. doi: 10.1207/s15327817la1103\_2
- Slabakova, R., Kempchinsky, P., and Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: a case of successful acquisition at the syntax–discourse interface. *Second Lang. Res.* 28, 319–343. doi: 10.1177/0267658312447612
- Sorace, A. (2011). Pinning down the concept of interface in bilingualism. *Linguist. Approac. Biling.* 1, 1–33. doi: 10.1075/lab.1.1.01sor
- Sorolla, N. (2009). “Casa, escola i pati: usos lingüístics quan s’acaben les classes. Síntesi i anàlisi de l’estudi sociodemogràfic i lingüístic als centres d’educació secundària de Catalunya,” in *NovesSL, Hivern 2009* (Barcelona: Generalitat de Catalunya).
- Suñer, M. (1995). Negative elements, Island effects and resumptive no. *Linguist. Rev.* 12, 233–273. doi: 10.1515/tlir.1995.12.3.233
- Tsimpli, I. (2014). Early, late or very late: timing acquisition and bilingualism. *Linguist. Approac. Biling.* 4, 283–313. doi: 10.1075/lab.4.3.01tsi
- Tubau, S. (2008). *Negative Concord in English and Romance: Syntax-Morphology Interface Conditions on the Expression of Negation*. Utrecht: LOT Publications.
- Unsworth, S. (2015). “Amount of exposure as a proxy for language dominance in bilingual acquisition,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán, and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173.
- Vallduví, E. (1994). Polarity items, n-words and minimizers in Catalan and Spanish. *Probus* 6, 263–294. doi: 10.1515/prbs.1994.6.2-3.263
- Villegas, Á. (2014). *The role of L2 English Immersion in the Processing of L1 Spanish Sentence Complement/Relative Clause Ambiguities*. State College, PA: Pennsylvania State University.
- von Heusinger, K. (2002). Specificity and definiteness in sentence and discourse structure. *J. Semant.* 19, 245–274. doi: 10.1093/jos/19.3.245
- von Heusinger, K., and Kaiser, G. (2003). “The interaction of animacy, definiteness and specificity in Spanish,” in *Proceedings of the Workshop on Semantic and Syntactic Aspects of Specificity in Romance Languages*, eds K. von Heusinger and G. Kaiser (Konstanz), 41–66.
- White, L. (2018). “Formal linguistics and second language acquisition,” in *Bilingual Cognition and Language: The State of the Science Across Its Subfields*. eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins), 57–78.
- Wulff, S., and Ellis, N. (2018). “Usage-based approaches to second language acquisition,” in *Bilingual Cognition and Language: The State of the science across its subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins), 37–56.
- Yilmaz, G., and Schmid, M. (2018). “First language attrition and bilingualism: adult speakers,” in *Bilingual Cognition and Language: The State of the Science Across Its Subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins Publishing Company), 225–250.
- Zanuttini, R. (1991). *Syntactic Properties of Sentential Negation*. Philadelphia, PA: University of Pennsylvania.
- Zeilstra, H. (2004). *Sentential Negation and Negative Concord*. Utrecht: LOT Publications.

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



# Predictors of Language Dominance: An Integrated Analysis of First Language Attrition and Second Language Acquisition in Late Bilinguals

Monika S. Schmid<sup>1\*</sup> and Gülsen Yilmaz<sup>2</sup>

<sup>1</sup> Centre for Research on Language Development throughout the Lifespan (LaDeLi), Department of Language and Linguistics, University of Essex, Colchester, United Kingdom, <sup>2</sup> Institut für Anglistik und Amerikanistik, Humboldt-Universität zu Berlin, Berlin, Germany

## OPEN ACCESS

### Edited by:

Dobrinka Genevska-Hanke,  
University of Oldenburg, Germany

### Reviewed by:

Adriana Belletti,  
Université de Genève, Switzerland  
Gloria Chamorro,  
University of Kent, United Kingdom

### \*Correspondence:

Monika S. Schmid  
mschmid@essex.ac.uk

### Specialty section:

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

**Received:** 04 April 2018

**Accepted:** 09 July 2018

**Published:** 20 August 2018

### Citation:

Schmid MS and Yilmaz G (2018)  
Predictors of Language Dominance:  
An Integrated Analysis of First  
Language Attrition and Second  
Language Acquisition in Late  
Bilinguals. *Front. Psychol.* 9:1306.  
doi: 10.3389/fpsyg.2018.01306

Late bilinguals who spend (part of) their adult lives in an environment where a language other than the one they learned in childhood is spoken typically experience a range of language development phenomena. Most obviously, they will acquire some level of receptive and/or productive knowledge of the new, or second, language (L2). How basic or advanced that level will be is determined by a range of environmental, experiential, attitudinal and individual factors. Secondly, they will most likely find the knowledge of their native language (L1) beginning to diverge from that of monolingual speakers in their country of origin, a process known as *language attrition*. In the course of this developmental process, some L2 skills may eventually match or even overtake the corresponding skill in the L1. This shift in the balance between L1 and L2 is the focus of investigations of *language dominance*. The present study explores language dominance in four migrant populations (Germans in the Netherlands and Canada, Turks and Moroccans in the Netherlands). Investigating both the development of formal/controlled skills and more automatic aspects of lexical access and fluency, we aim to attain an understanding of how extralinguistic factors contribute to the development of *both* languages. We argue that an integrated perspective can contribute more profound insights into the predictors of this complex process of bilingual development. In particular, our findings show that statistical models based on linear relationships fall short of capturing the full picture. We propose an alternative method of analysing data, namely discriminant function analysis, based on a categorisation of the populations, and demonstrate how this can enhance our understanding. Our findings suggest that different aspects of the bilingual experience contribute differently to language development, regardless of language combination and type of skill measured. Contrary to what previous research suggests, measures relating to the intensity of informal use of both the L1 and the L2 in daily life are important in determining whether someone is a good or a poor L1 maintainer, while high vs. low success in acquisition appears to be predominantly associated with personal factors such as educational level.

**Keywords:** bilingual development, language attrition, second language development, late bilinguals, language dominance, language balance, extralinguistic factors, non-linear statistical models

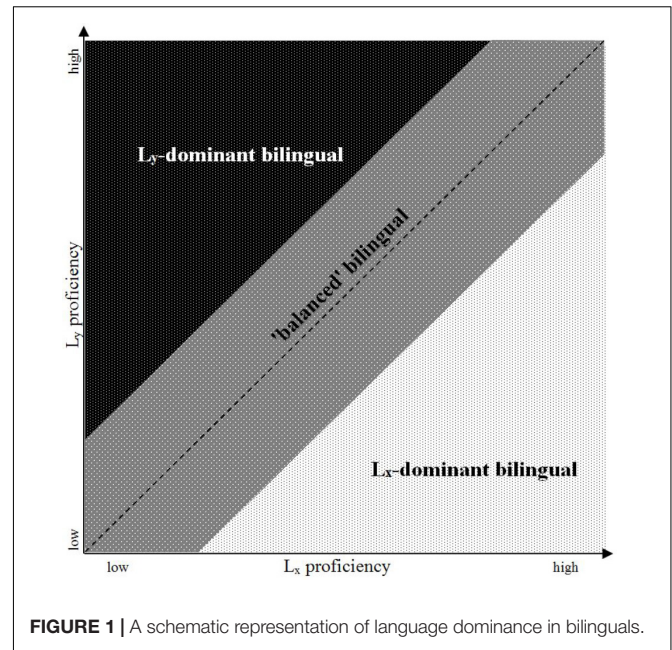
## INTRODUCTION

Language dominance is an extremely complex concept, encompassing a wide range of aspects and features (e.g., Gertken et al., 2014; Silva-Corvalan and Treffers-Daller, 2016). These features can roughly be divided into two sets: the first set consists of those aspects of language that usually constitute the outcome measures or dependent variables for linguistic investigations – that is, measurable phenomena that relate to the knowledge, use and processing of all of a bilingual’s languages at all linguistic levels – and fall under the broad concept of ‘proficiency.’ The second set comprises measures related to personal background variables such as age, education, or language aptitude; the context in which the languages were acquired; language experience and habits; and linguistic and cultural identification. These factors are usually the independent variables, they predict the extent to which the first set of variables is developed in any one individual speaker. Language dominance, therefore, “takes into account the two languages of a bilingual person, not just one, biographical variables and the language-external conditions under which the two languages are learned or used by bilinguals” (Montrul, 2016, p. 17).

In terms of the first factor set, which for the sake of simplicity we will refer to here as ‘proficiency measures,’ every bilingual speaker can therefore be situated somewhere in a two-dimensional space defined by an  $x$ -axis representing language  $X$  ( $L_X$ ) and a  $y$ -axis, representing language  $Y$  ( $L_Y$ ) (see **Figure 1**)<sup>1</sup>. A speaker who is mapped close to the diagonal of this space (that is, at a similar level on both axes) is someone whose proficiency is more or less ‘balanced’ between the two languages, while one who is closer to one axis than the other is ‘dominant’ in the ‘stronger’ language, the one in which s/he has scored more highly. It should be noted that this visualisation is a simplified and idealised one: ‘proficiency’ cannot easily be reduced to a single measure (see section “Outcome Variables in Bilingual Development: Definitions and Measurements of ‘Proficiency’” below), and the position of the same individual may therefore vary considerably depending on what skill or task is being measured (e.g., Bahrack et al., 1994; Kupisch and van de Weijer, 2016; Montrul, 2016).

The predictor variables are equally problematic in terms of definition and measurement, as this factor set covers a wide range of aspects of the bilingual experience and is therefore as varied as are the bilingual individuals themselves. Various models attempt to capture the multi-facetedness and multi-dimensionality of language exposure and use, for example the *Complementarity Principle* which assigns each domain of use – such as the home, politics, specific leisure activities etc. – one or several languages associated with it in a particular bilingual individual’s experience, reflecting the recognition that “[d]ifferent aspects of life require different languages” (Grosjean, 1997, p. 165; see also Grosjean, 2016). A ratio-based calculation (similar to the one used to establish handedness), based on absolute proficiency in both

<sup>1</sup>The number of dimensions can be extended to encompass more than two languages. The discussion here will be confined to a two-language scenario for the sake of simplicity, but may equally be taken to refer to multilinguals proficient in a potentially unlimited number of languages.



**FIGURE 1** | A schematic representation of language dominance in bilinguals.

languages and capable of capturing language dominance as a gradient phenomenon, is described in Birdsong’s (2016) proposal of a *Dominance Index*.

The third dimension in the imaginary space defined by proficiency in  $L_X$  and  $L_Y$  in **Figure 1** is *time*: development unfolds as the independent variables exert their influence on the dependent ones (e.g., when an increase in exposure to  $L_Y$  leads to a higher level of proficiency in that language, potentially also affecting  $L_X$ ). Time, in this model, is the only dimension which is linear and unidirectional. Linguistic development is neither: it can shift toward the higher or the lower end of the spectrum in either language, encompassing both acquisition and loss. Such shifts can occur in bursts or slowly and gradually, they can reverse direction from growth to decline and back again and they can affect mainly one language, both languages equally, or both languages orthogonally (a growth in one language occurring alongside a decline in the other).

The formidable task for investigations of language dominance, then, is to provide explanatory models capable of mapping out how the predictor variables may interact and determine the intensity and direction of developmental changes for any given aspect of language proficiency over time. Crucially, we argue here that it is important to take into account *both* linguistic dimensions in order for such models to fully capture the phenomenon and not to reduce the analysis to the development of one language only, nor to collapse them into a one-dimensional function, e.g., by subtracting one from the other.

## INVESTIGATIONS OF DEVELOPMENT IN BOTH LANGUAGES

To date, such integrated models of bilingual development have been strongly biased toward speakers for whom development



in both languages goes hand in hand: simultaneous or early bilinguals for whom acquisition of  $L_y$  begins while  $L_x$  is still at an early stage of development. Findings indicate that the highly active phases of language development during childhood and adolescence allow great flexibility when it comes to shifts in language proficiency and dominance, with changes in external circumstances or exposure – such as the start of (nursery) school, a move between countries, or a return to the home country – often causing spurts of development and/or regression that may fundamentally change the overall multilingual balance within months (e.g., Flores, 2015; de Houwer and Bornstein, 2016).

In late bilingualism, on the other hand, the most intensive stages of development in each of an individual's languages take place during different life phases, with the onset of second language acquisition (SLA) occurring after the development of the L1 has reached a relatively mature level and L1 development has hence slowed down considerably. That being the case, the majority of investigations of late bilingual development focus on the L2, the assumption being that a level of stability – some kind of 'steady state' – has been reached in the L1 which makes it uninteresting for research (e.g., Gregg, 2010). This notion has been challenged in the context of research on *first language attrition* (L1At) which argues that the addition of a new language will inevitably lead to changes in all of those aspects that we subsume here under the term *proficiency* in the language that is already established. A growing body of research provides evidence of such attrition in immersed as well as non-immersed bilinguals (e.g., Schmid and Köpke, 2017) and attempts to probe the relationship between predictor and outcome variables in this process (e.g., Schmid and Dusseldorp, 2010).

An immediate question in this respect is to what extent a gain with respect to any particular aspect of proficiency in one of a bilinguals' languages may be associated with (or even cause) a loss in the other. When it comes to simultaneous or early bilingualism, popular understanding often has it that Language X will grow at the expense or to the detriment of Language Y. Research on bilingual development has long since demonstrated that there is no such straightforward or trivial relationship (e.g., Cummins, 1991; Bialystok, 2001). Bilingual children may develop both of their languages at the same pace, hitting the milestones assumed for typically developing monolinguals at roughly the same age in both, or development may be quite asymmetrical, favouring one language over the other at different stages. A developing bilingual child may occupy any position in the imaginary space mapped out in **Figure 1** above.

When it comes to L1At, findings on the relationship between proficiency in the speaker's two languages are quite sketchy. In early research a relatively straightforward correlation was often assumed, according to which "[t]he greater the degree to which a speaker masters one system, the greater the extent to which one might expect it to affect another system," so that "we expect L1 loss to be greatest in individuals who [...] have mastered the L2 to a relatively high degree" (Major, 1992, p. 191). Opitz (2011) ascribes this hypothesised 'trade-off' in proficiency to the languages "competing for potentially insufficient resources required for maintaining the languages simultaneously at the desired high level" (p. 81). Other studies,

however, have hypothesised that the development of proficiency in both L1 and L2 may be less straightforward, highly task-dependent, and modulated by other factors – for example, a high level of language aptitude may allow a particular speaker to acquire and maintain high proficiency levels in both languages (e.g., Cummins, 1991; Bylund et al., 2009; Cherciov, 2013).

There are few studies to date which investigate proficiency in both the L1 and the L2 of late immersed bilinguals. The findings that do attempt to make direct comparisons across languages and tasks (e.g., Dostert, 2009; Opitz, 2010, 2011, 2013; Cherciov, 2011) or link the amount of L1 attrition to the level of proficiency in the L2 (e.g., Baladzhaeva, 2013) seem to suggest that, similar to the early bilinguals discussed above, L1 attriters can fall anywhere on the spectrum: while in some cases (weak) correlations are observed between measures in both languages, a look at individual data suggests that this relationship is anything but deterministic. In other words, some participants attain high proficiency in the L2 but perform poorly in the L1, for others it is the other way around, while yet others are extremely good or quite poor in both their languages. This poses a problem for empirical investigations, since it implies that establishing language dominance based on difference scores (as is extremely common, e.g., the overview in Treffers-Daller, 2016) falls short of capturing the full picture. Such an approach ranks someone who does extremely well in both languages (a good maintainer and good learner) exactly the same as someone who performs very poorly across the board (a poor maintainer and poor learner) (see also Birdsong, 2016).

Treffers-Daller (2016, p. 261) takes this problem into account in her 'typology of language dominance based on language proficiency' which, instead of dividing the proficiency spectrum shown in **Figure 1** above into three sections ( $L_x$ -dominant,  $L_y$ -dominant and balanced) uses four quadrants: dominant bilinguals (either  $L_x < L_y$  or  $L_x > L_y$ ), low-achieving balanced bilinguals and high-achieving balanced bilinguals. Treffers-Daller (2016, p. 262) acknowledges, however, that this typology lacks explanatory value, since it "does not indicate how proficiency in these languages has developed". She argues that future research agendas should therefore focus on the interaction between language use and ability in order to address this knowledge gap. A qualitative approach to doing so is suggested by Opitz (2016, in press) through scrutinising particularly good exemplars of the four developmental types. While such an approach is useful for gaining preliminary insights into the developmental processes and the factors which drive them, we propose that avenues should also be explored which allow for the quantitative/experimental exploration of large datasets in order to empirically verify such observations. The present study is a tentative attempt at one such approach.

## PROBLEMS OF MEASUREMENT

### Predictor Variables in Bilingual Development

Where second language development is concerned, quality and quantity of *input* and *output* are among the most important



predictors (e.g., Gass and Mackey, 2007) with a number of modulating variables linked to factors that are usually referred to as *individual differences*, such as motivation, aptitude and cognitive style (e.g., Dörnyei, 2005). The picture is much more obscure when it comes to the impact of predictor variables on L1At (e.g., Opitz, in press). A number of hypotheses have been advanced in this context, the most common one being that a lack of exposure to and use of the L1 will lead to its deterioration (e.g., Cook, 2003) – in other words, that the amount of L1At to be observed will correlate negatively with the degree of use of that language in daily life, and that these effects will increase with a longer period of residence in the host country. This hypothesis echoes the relationship between use and success observed in SLA. Secondly, and also in line with what has been found for SLA, it has often been predicted that attrition will be modulated by attitude and motivation, a more positive attitude toward the language itself and the speech community facilitating language maintenance (e.g., Cherciov, 2013). While both of these predictions may appear self-evident, empirical research to date has been strikingly unsuccessful in substantiating any such relationships (see Cherciov, 2011, 2013; Schmid, forthcoming for reviews) suggesting that, if there is an impact at all of exposure/use and attitude/motivation on L1At, it is much less pronounced and/or more complex than it is in SLA. In particular, while re-immersion in a monolingual context appears to help regain native-like levels on some aspects of proficiency that were shown to be attrited prior to the re-exposure (e.g., Chamorro et al., 2016; Genevskaya-Hanke, 2017) L1 use in daily life in the immigration setting has not been shown to be systematically related to performance (Schmid, forthcoming).

This difficulty of empirically establishing a connexion between predictor and outcome variables in the process of L1At has sometimes been ascribed to the assumption that the influence of neither individual predictors nor their interaction may produce an effect that is linear (e.g., de Bot et al., 1991; Schmid, 2011b) and that it may thus elude capture by means of traditional statistical techniques. A number of inevitable practical and methodological considerations further complicate matters: the multi-facetedness of the bilingual experience – particularly in the immersion setting typical for attrition studies – necessitates that the research design should include a large number of independent variables (see Schmid, 2011a). Many of these have to rely on self-assessments, usually elicited by means of Likert-scale type questions, and they are almost invariably non-normally distributed. For example, within the much-studied communities of the traditional ‘guest worker’ immigrants who arrived in Western European countries in the 1960s and 70s due to labour shortages, few speakers will report that they often use their native language for professional purposes, so this variable is likely to be skewed toward the L2 for such populations. In general it is, in our experience, quite rare for individual speakers to state that they use both languages equally in any one domain. This is a natural consequence of Grosjean’s Complementarity Principle (see above) – bilinguals use different languages to do different things – but it implies that most predictors will either be skewed toward one or the other language or show a bimodal distribution within communities that are less homogenous. While such problems of multifacetedness,

non-linearity of interactions and non-normality of distributions have to be acknowledged as probably inevitable complicating considerations for language attrition research, we propose here that they can be dealt with by choosing the appropriate statistical procedures.

Beyond such methodological and practical considerations, however, we suggest that the lack of insight may be related to the fact that, to date, most of the empirical and quantitative work on L1At has limited itself to investigating the L1 (see Opitz, 2013), while studies of L2 development tend to only be interested in the L2, assuming the L1 to be an invariate baseline. The present study seeks to investigate to what extent our understanding can benefit from an approach that investigates language proficiency as a fully two-dimensional construct. This begs the question of what we understand by ‘language proficiency.’

## Outcome Variables in Bilingual Development: Definitions and Measurements of ‘Proficiency’

The problems and considerations listed above represent a formidable set of challenges for investigations of language dominance – they pale, however, in comparison with the difficulties involved in defining and measuring the elusive notion of ‘language proficiency.’ This is a catch-all term that has been used to describe radically different aspects of language skills and measurements, depending on the population under investigation and the theoretical framework within which a study is conducted (see Hulstijn, 2012 for an overview). For example, proficiency has been operationalised as mean length of utterance (MLU) in investigations of child language development (e.g., Yip and Matthews, 2006), as the ability to fill in the gaps in a cloze test in instructed second language learning (Tremblay, 2011), as the ability to pass as a native speaker in investigations of maturational limits to ultimate attainment (e.g., Bongaerts et al., 1997), as the ability to name objects on a computer screen quickly and accurately in investigations of language processing (e.g., Mägiste, 1992), as the ability to use the language with native-like levels of fluency (e.g., de Jong et al., 2015) or as the ability to recognise and process violations of particular grammatical features in studies of the development of underlying mental grammars and representations (e.g., Hopp, 2007) – to name but a few.

Studies of language dominance, which attempt to assess relative levels of proficiency across languages, first of all have to acknowledge that some of the skills or measures listed above lend themselves more readily to direct crosslinguistic comparisons of  $L_x$  and  $L_y$  development (for example, MLU or naming latencies are often used to establish levels of language dominance) while for others it is more difficult to see how crosslinguistic equivalence can be established (see Hulstijn, 2012; Montrul, 2016; Treffers-Daller, 2016 for discussion). In particular, specific questions of grammatical development (based, for example, on theoretical issues concerning parametrisation, interfaces etc.) are hard to address in such a framework, as by their very nature they will focus on features which are hard to acquire or maintain in only one of a bilingual’s languages, making meaningful comparisons across both dimensions hard.

For investigations which aim to provide an integrated model of (global) proficiency in  $L_x$  and  $L_y$  among late bilinguals, methods which focus on tasks or measurements related to particular aspects of the structure of each language (such as grammaticality judgments or the production/perception of certain phonemes) are therefore problematic. Instead, the outcome variables should be selected to represent relatively general and holistic aspects of language proficiency, taking into account the extent to which these may vary crosslinguistically in both native and non-native populations.

Given that L1At populations command two languages which are learned under similar conditions (naturalistic learning through immersion in the linguistic community) but at different stages in life – that is, speakers who unambiguously have one native and one non-native language – a fruitful framework for the assessment of proficiency is the model proposed by Hulstijn (2011, 2015) which distinguishes Shared/Basic Language Cognition (BLC) and Extended/Higher Language Cognition (HLC). In this model, BLC refers to

- (a) the largely implicit, unconscious knowledge in the domains of phonetics, prosody, phonology, morphology and syntax;
- (b) the largely explicit, conscious knowledge in the lexical domain (form-meaning mappings), *in combination with* (c) the automaticity with which these types of knowledge can be processed (Hulstijn, 2011, p. 230, his emphasis)

and is restricted to those lexical items and morphosyntactic structures in spoken language which all adult L1 speakers share (irrespective of their age, educational level, or level of literacy) and which they may encounter in all communicative situations. HLC, on the other hand, refers to more complex domains of use, encompassing less frequent items and structures as well as written language, and it is in this domain that native as well as non-native speakers vary considerably from each other. ‘Frequency,’ in this framework, is operationalised on the basis of the assumption that lexical items and grammatical structures follow a Zipfian distribution in naturally occurring language, where ‘highly frequent’ items (those belonging to BLC) are situated on the steep left side of the slope, while less frequent items are to be found on the flattening part of the curve to the right (Hulstijn, 2015:22ff.)

Among the fundamental assumptions of the model are

- (a) Command of BLC should be homogenous within as well as across native populations, while these populations will be stratified when it comes to HLC.
- (b) Non-native populations will exhibit considerable variability with respect to both; they are unlikely to fully acquire BLC in their L2 beyond the domains of vocabulary but may nonetheless reach high native levels of HLC
- (c) Development of BLC in non-natives and HLC in both populations will be modulated by a combination of external and background factors (Hulstijn, 2015:52f.)

Hulstijn’s model thus assumes two types of speakers: the ‘native speaker,’ who will be at ceiling for all components of BLC

but may vary with respect to HLC, and the ‘non-native speaker’<sup>2</sup> who will exhibit variability in both BLC and HLC. Attriting populations, however, are similar to late L2ers in that they diverge from monolingual control populations in both domains in their L1, indicating that even for native speakers, becoming bilingual will affect performance on skills belonging to both BLC and HLC.

### Basic Language Cognition in L1 Attrition

There are a host of findings demonstrating that many of the language components which belong to BLC, and which are therefore assumed to show little variance among ‘native speakers,’ are subject to change and L1 attrition in immersed late bilinguals. These include<sup>3</sup>:

**Accentedness:** while monolingual populations are typically perceived to be at ceiling in global foreign accent rating experiments, several studies have established an increase in variance of such ratings in immersed late bilinguals which can lead to some attriters being perceived as unambiguously non-native (e.g., de Leeuw et al., 2010; Hopp and Schmid, 2013; Bergmann et al., 2016; Karayayla, 2018: ch. 4) and subtle shifts occurring in the production of both segmentals and suprasegmentals away from monolingual native norms (e.g., Mennen, 2004; de Leeuw, 2008; Chang, 2012; Bergmann et al., 2016).

**Fluency:** in both free and elicited discourse, L1At populations have consistently been demonstrated to be less fluent than monolinguals, as indicated by a slower speech rate and higher incidence of pauses, filled pauses, repetitions and self-corrections (e.g., Dostert, 2009; Schmid and Fägersten, 2010; Cherciov, 2011; Yilmaz and Schmid, 2012; Bergmann et al., 2015).

**Lexical access:** L1At populations are less productive at generating lexical items in Verbal Fluency tasks (e.g., Waas, 1996; Yagmur, 1997; Keijzer, 2007; Varga, 2012; Schmid and Jarvis, 2014) and slower and less accurate in naming tasks (e.g., Mägiste, 1992; Ammerlaan, 1996; Baus et al., 2008) than monolingual controls, suggesting that their access to even the highly frequent elements that are typically elicited in such tasks is delayed.

**Overt/null pronouns:** This is an example of grammatical features which form part of BLC. There are a large number of investigations demonstrating that overt pronouns come to be overgeneralised to contexts where monolingual natives would

<sup>2</sup>The model further identifies subpopulations of ‘non-natives’: (a) early bilinguals and heritage speakers, (b) immersed second language learners and (c) instructed second language learners (Hulstijn, 2015: 47f.). For the purpose of the present study discussion will be restricted to Category B – those speakers who typically form the populations under investigation in attrition studies.

<sup>3</sup>A full review of findings on L1At being beyond the scope of this paper we refer the reader to Schmid and Köpke (2017). We limit the features we discuss to relatively global measures that can be compared between a bilinguals’ two languages and exclude more language-specific features, such as tense and aspect (e.g., Montrul, 2008) or number agreement (Kasparian et al., 2017).

use null pronouns by attriters of pro-drop languages such as Bulgarian, Italian, Greek, and Spanish (e.g., Tsimpli et al., 2004; Domínguez, 2013; Genevskaya-Hanke, 2017).

### Higher Language Cognition in L1 Attrition

Given that even monolingual native populations are assumed to be stratified with respect to HLC, it is hardly surprising that this variance increases under the cognitive demands of bilingualism. A wide range of studies have demonstrated this, for example, for complex and infrequent syntactic phenomena. For example, several studies of embedding structures in L1 Turkish have found attriters to diverge most from monolinguals on those types of embedding which are morphologically the most complex (i.e., involve the highest number of suffixations/transformations) and occur least frequently in free speech (e.g., Yagmur, 1997; Yılmaz, 2011; Karayayla, 2018). In a similar vein, attriters are consistently outperformed by non-attrited controls when it comes to the completion of complex written tasks (such as C-tests or cloze tests, e.g., Schmid and Dusseldorp, 2010; Cherciov, 2011; Varga, 2012; Kasparian, 2015).

In addition to distinguishing between BLC and HLC, Hulstijn's model furthermore differentiates 'core' and 'periphery' aspects of language proficiency: core components refer largely to linguistic knowledge and the speed with which it can be processed, while peripheral components refer to the more metalinguistic skills, such as interactional ability and knowledge of the characteristics of different types of (spoken or written) discourse (Hulstijn, 2015, p. 41). The development of these HLC/peripheral skills in L1At has often been assessed based on the Can-Do Scales developed within the Common European Framework of Reference (see Hulstijn, 2015: ch. 10 for an in-depth discussion of the CEFR and its relationship to the BLC/HLC model, see below section "Study 1: Proficiency Measures Relating to Higher Language Cognition (HLC)" for details on the scales), and such studies tend to find larger differences between attriters and controls when it comes to reading and writing skills than with respect to speaking and listening (e.g., Opitz, 2011). This finding tentatively suggests that HLC components of language proficiency may be more vulnerable than BLC components – a hypothesis in need of further corroboration, but in line with the basic assumptions of the model.

### Bilingual Development and the BLC/HLC Model

The findings presented above suggest that the BLC/HLC model can profitably be extended to include the development of language proficiency under conditions of L1At. The fact that the model assumes a global and holistic approach to defining language proficiency furthermore makes it ideally suited for an investigation of the relative development of proficiency in both the L1 and the L2 within the framework of language dominance, as the components of the model can be assessed and compared across languages. The findings presented above illustrate that, for linguistic features in both domains, attriting populations develop increased variability and diverge from the native baseline.

What remains entirely unclear, however, are the conditions or predictors which drive these changes: In all of the studies listed above, some of the attriters remain within the native range while others fall squarely outside it, but assessments of the impact of predictor variables have remained largely inconclusive (e.g., Schmid, 2011b). In other words, it is unclear to what extent external factors such as the frequency and domains of L1 and L2 use, the length of residence, or levels of attitude and motivation, contribute to the deterioration or maintenance of any particular linguistic feature. This paper attempts to address this knowledge gap by adopting an innovative approach that we believe is capable of assessing the development in both languages within an integrated framework. We will focus on those domains of language that allow us to make meaningful comparisons of the level of development in L1 and L2, namely measures related to lexical access and to the level of ability of performing tasks related to written language.

## HYPOTHESES AND RESEARCH QUESTIONS

The present study sets out to test the hypothesis that the explanatory potential of investigations of language dominance can be enhanced by adopting a fully two-dimensional approach which takes into account performance in both L1 and L2. We furthermore assume that linear statistical models – that is, models based on regression slopes – may not be able to capture the complex interaction of different features of proficiency, personal background, exposure/use, and attitudes (henceforth: external factors), and propose that a classification into different types of language developers may allow a more detailed picture to emerge.

We ask the following research questions:

- RQ1: What is the predictive power of external factors for measures of language proficiency in both the L1 and the L2 of long-term immersed bilingual populations in a multifactorial linear model?
- RQ2: To what extent can the predictive power of external factors for measures of language proficiency in both the L1 and the L2 of long-term immersed bilingual populations be increased by adopting a data categorisation approach which does not rely on the linearity of relations?
- RQ3: What are the differences between HLC and BLC aspects of proficiency with respect to RQs 1 and 2?

## THE STUDY

### Ethical Approval

The data reported here were collected in 2004 (Study 1) when the PI (the first author of this paper) was affiliated with the Vrije Universiteit Amsterdam and in 2007 (Study 2) when both authors were affiliated with the University of Groningen. At this time, the humanities faculties at these institutions did not have a protocol for ethical approval nor an ethics committee, and there were no national guidelines in relation to this. All



participants did provide written informed consent prior to the experiment. With hindsight we recognise the lack of formal ethical approval for the studies to be a shortcoming – which, unfortunately, cannot be addressed retrospectively. However, all of the materials and experiments reported on here have been used by both authors in subsequent investigations for which ethical approval was duly granted according to the protocols required by different institutions, including the University of Groningen, the University of Essex and the Humboldt University, Berlin. We are therefore convinced that the research design in itself is unproblematic from an ethical point of view.

## Participants and Predictor Variables

The data for the present study were collected from four experimental and three control populations. Study 1, which focuses on the development of Higher Language Cognition (HLC), was conducted with native German speakers ( $n = 106$ ) with between 9 and 58 years of residence (LoR) in the Netherlands ( $n = 53$ , mean LoR 34.28) and the Greater Vancouver area, Canada ( $n = 53$ , mean LoR 37.09). Study 2, investigating aspects of Basic Language Cognition (BLC), was conducted with 87 migrants with between 10 and 43 years of residence in the Netherlands. 52 of these speakers were Turkish natives (mean LoR 22.57), while 35 were native speakers of Moroccan Arabic (mean LoR 23.31).

The experiments described hereunder investigate both the L1 and the L2 of these speakers. In Study 1, data collection was done in a single session, as the collection of L2 data was restricted to two tasks tapping into controlled and highly monitored language skills. Study 2, on the other hand, was conducted in two different sessions and by different researchers, due to considerations linked to language mode: Session 1 collected both experimental and informal spoken L1 data, while Session 2 (which took place several months later) collected similar data in the L2. In order to induce a predominantly monolingual language mode for these experiments, we considered it important that the researcher should be a speaker of the language which was the focus of the experimental session (Turkish/Moroccan Arabic in Session 1, Dutch in Session 2) with no knowledge of the other language. The researchers conducting Session 1 were recent arrivals to the Netherlands and native speakers of Turkish and Moroccan Arabic, respectively. Session 2 was conducted by two research assistants who were native speakers of Dutch but had no knowledge of either Turkish or any variety of Arabic. Unfortunately but inevitably this led to some participant loss, with data from only 63 participants available at Session 2. In this and the following sections, the dataset comprising 87 participants will be referred to as ‘Full dataset,’ while the dataset comprising the 63 participants with L2 data available will be referred to as ‘Limited dataset.’ **Table 1** summarises participant characteristics for both studies.

## Language- and Attitude-Related Background Factors

Data on participants’ biography, language learning history, language use and language attitudes were collected by means

of the same questionnaire in both studies<sup>4</sup>. The questionnaire comprises a total of 77 questions in different formats: open questions (e.g., birthplace, profession, personal reflections), Likert-scale questions (e.g., levels of use, attitudes and preferences), and interval questions (e.g., age). The questionnaire and its coding and analysis are described in detail in Schmid (2011a). The questionnaire was used by the researcher as the basis for a semi-structured interview, where the participants were prompted to talk about themselves, their biography and their languages, freely, informally and in detail (the procedure for conducting such an interview is described in Schmid, 2011a). All interviews were transcribed and coding was checked against both the recording and the notes taken during the session. The variables derived from this and used in the present study are described below, an overview of responses per question and group is presented in the **Supplementary Table S1**.

### *Self-reported language proficiency*

All participants were asked to rate their proficiency in both their L1 and L2 (used here to refer to the language of the country in which they were living at the time of data collection) both at the time of migration and at the time of the interview, and also to state which of these languages they felt was the stronger at the present time. There were several interesting differences between groups, such that regardless of L1 background almost none of the migrants to the Netherlands knew more than a few words of Dutch before arrival while more than half of the migrants to Canada rated themselves as intermediate or proficient in English at arrival. At the time of testing, most of the bilinguals felt they had intermediate or good proficiency in the L2, with the English L2 speakers again standing out. The Germans rate their proficiency in Dutch more highly than the Turks and the Moroccans, possibly reflecting the advantage the close typological relatedness between their L1 and their L2 gives them. With very few exceptions, everyone rated their L1 proficiency at migration as ‘good’ or ‘very good,’ but that proportion dropped across the board for proficiency at the time of testing, although only one single speaker described it as ‘bad.’ Only among the L2 English speakers did more than a quarter of participants feel that their L2 had become stronger than their L1, while just over half of all L1 Germans thought both languages were equally good. Balanced bilingualism or dominance reversal were much rarer among the Turks and the Moroccans, with strong majorities in both groups feeling that their L1 remained their stronger language.

### *Language exposure and use*

The questionnaire contains a total of 25 5-point Likert Scale questions on frequency of L1 exposure and use:

- L1 and L2 use within the family (with partner – 4 items, with children – 4 items, with grandchildren – 4 items)
- L1 and L2 use with friends and acquaintances (4 questions)
- L1 and L2 use at work (2 questions)
- Frequency of use of L1 media (radio, tv, newsmedia, books, 4 questions)

<sup>4</sup><https://languageattrition.org/resources-for-researchers/experiment-materials/sociolinguistic-questionnaires/>



**TABLE 1** | Participant characteristics.

	Study 1		Study 2 (full dataset)		Study 2 (limited dataset)	
L1	German	German	Turkish	Moroccan Arabic	Turkish	Moroccan Arabic
L2	English	Dutch	Dutch	Dutch	Dutch	Dutch
<i>n</i>	53	53	52	35	46	17
Females <i>n</i> (%)	34 (61.2)	35 (66.0)	32 (61.5)	7 (20)	31 (67.4)	7 (21.5)
Age (range)	63.23 (37–88)	63.36 (37–85)	43.15 (28–61)	46.63 (30–65)	43.04 (28–61)	43.57 (31–65)
Age at emigration (range)	26.13 (14–47)	29.08 (16–51)	20.38 (14–42)	24.06 (18–32)	19.74 (14–42)	24.29 (18–32)
Length of residence (range)	37.09 (9–54)	34.28 (14–58)	22.15 (10–35)	22.57 (10–43)	22.62 (10–35)	19.29 (10–38)
Educational level (%)	Primary	13 (24.5)	9 (17.0)	17 (32.7)	10 (28.6)	16 (34.0)
	Secondary	27 (50.9)	27 (50.9)	10 (19.2)	22 (62.9)	8 (17.0)
	Tertiary	13 (24.5)	17 (32.1)	25 (48.1)	3 (8.6)	23 (48.9)

- Frequency of visits to the home country (1 question)
- Overall contact with and use of the L1 (2 questions)

Some of these items had to be excluded for the present analysis as variability was too low (for example, virtually all of the Netherlands-based participants stated that they visited their home country at least once a year). The general picture which emerged across these questions was that most participants continued to use their L1 on a fairly regular basis, the Germans slightly less so than the Turks/Moroccans. All groups had good social contacts within their new country but there was some variance, with roughly two-thirds of Turks and Moroccans reporting more friends who shared their L1 while over half of the Germans said their social network was composed mainly of native speakers of the L2. When it comes to language use in the family, the Turks stand out somewhat from the other groups with a much stronger claimed adherence to an L1-only policy, while half of the Germans report using L2 only (many people noted in the interview that they would have liked to have persisted more on using their L1 with their children, but that they had faced too much resistance and had given up). The Moroccans appear to occupy an intermediate position. A similar picture appears across most language exposure and use questions: most participants appear to have a fairly clear preference for one language in each context, with the Turks and Moroccans leaning more toward the L1 than the Germans. The only exception to this is the use of the native language for professional purposes, which stands at around 20% for both German groups but is quite rare for the Moroccans and Turks.

### Attitudes

Where attitudes toward the native language are concerned, the views seem more homogenous across groups, with over 75% in all groups saying it is important or very important to them to maintain their L1 and almost the same proportion of respondents saying it is important to them to pass it on to their children. An interesting finding emerged from the question “Which language do you prefer?”: while the Germans in Canada were split roughly evenly between their L1 and their L2, and three quarters of the Turks and Moroccans stated a preference for their L1, the only group that had a substantial proportion of self-reported balanced bilinguals for this questions

(that is, of speakers who report ‘no preference’) were the Germans in the Netherlands. This suggests that the similarity between L1 and L2 for the German–Dutch bilinguals may have facilitated the perception of a more balanced bilingualism. Interestingly, while there was only one speaker who reported ‘no preference’ on this question in the German–Canadian group, over 50% responded in the affirmative when asked whether they felt they were balanced bilinguals, while for all other groups, the answers across both questions seemed to be largely consistent.

### Principal component analysis

The overview of findings presented here points to two general problems concerning personal background data in language attrition research. Firstly, there are many questions with potentially important information for which there is missing data from a substantial proportion of informants – for example, not all participants have a partner and/or children. Secondly, as pointed out above (see section “Predictor Variables in Bilingual Development”), the data are not normally distributed: for most of the variables reported here, there is either a skewness toward the L1 end of the scale or a bimodal distribution. Both phenomena are a natural and inevitable characteristic of attrited populations: most studies find sustained preference for the native language, particularly where it comes to self-assessed proficiency. Furthermore, as is predicted by the Complementarity Principle (Grosjean, 1997, 2016), most people tend to prefer one or the other language across most domains.

In order to alleviate these problems as well as reduce the number of predictor variables for analysis, we conducted a Principal Component Analysis (PCA). We addressed the problem of missing values by replacing them in each case with the neutral point on the scale. We chose this strategy over the more common approach of imputing missing values based on the rationale that, for the data in question, values either above or below the neutral measure would (incorrectly) suggest that the relevant language (the L2 if the imputed value was below the mean and the L1 if it was above it) played a role in the prediction of the outcome variables while setting it to neutral allowed the case to be included in the analyses without such an effect.

**TABLE 2** | Outcome variables, German L1 group.

	Germans in Canada (GECA)		Germans in the Netherlands (GENL)		German controls (GECG)	
	Mean (std)	Range	Mean (std)	Range	Mean (std)	Range
C-Test L1	75.26 (11.61)	46–95	77.21 (13.81)	38–95	82.21 (8.90)	59–99
Can-Do L1	3.92 (0.59)	2.63–4.67	4.12 (0.46)	2.71–4.67	3.83 (0.5)	2.76–4.65
C-Test L2 original scores	70.42 (17.82)	15–92	76.31 (17.33)	12–99		
C-Test L2 standardised	71.98 (23.14)	0–100	73.92 (19.92)	0–100		
Can-Do L2	3.99 (0.61)	2.42–4.67	3.81 (0.68)	1.80–4.78		

All variables were standardised to the same scale prior to entry into the PCA, with the maximum value in the dataset (e.g., 88 in the case of *age at testing*) set to 1 and the minimum value set to 0. The PCA (Varimax Rotation, extraction of factors with *Eigenvalues* > 1.000) identified a total of six components which were saved as factors (see **Supplementary Table S2** for the full component matrix). The first component comprised 9 factors relating largely to the frequency of casual and informal use of the L1 and the L2, that is, with family and friends (Cronbach's  $\alpha = 0.881$ ) and was labelled *Interactive Use*. The second component, *Personal Background*, comprised the variables age, length of residence and education ( $\alpha = 0.609$ ). The third component related to *Perception* and comprised the answers to the questions about current L1 proficiency and whether that had changed since immigration ( $\alpha = 0.455$ ). Component 4 comprised the *Attitude*-related variables of importance to maintain the L1, transmit it to the children, and culture of preference, alongside the frequency of use of L1 media (books, TV, and radio) ( $\alpha = 0.627$ ). Overall *Contact* with the L1 was a unifactorial component, while *Professional Use* of both L1 and L2 made up the last component ( $\alpha = 0.476$ ).

The first four components were normally distributed. There was a slight negative skew for *Contact* [ $D(190) = 0.091$ ,  $p = 0.001$ ] and a more pronounced positive one for *Professional Use* [ $D(190) = 0.116$ ,  $p < 0.001$ ]. These variables were log-transformed after a constant was added to make all values positive and the scale for the negatively skewed component was inverted. The transformed components were no longer skewed [*Contact*:  $D(190) = 0.049$ ,  $p = 0.2$ ; *Professional Use*:  $D(190) = 0.061$ ,  $p = 0.086$ ].

## Outcome Variables

The data collection for both studies included a native language control group for the L1 tasks (German in Study 1, Turkish and Moroccan Arabic in Study 2), and Study 2 also used a Dutch native control group for the L2 tasks. Controls were matched with the relevant experimental populations for age, gender, educational background and, in Study 2, region of origin within the L1 country. It is not the purpose of the present investigation to probe into issues of general proficiency or overall attainment against an idealised monolingual baseline, but to assess to what degree development has taken place in both languages within the proficiency space defined by the performance of the immersed bilingual population. While the descriptive statistics given below include the results from the

control group as indicative values, they were therefore not used in the inferential statistics.

## Study 1: proficiency measures relating to higher language cognition (HLC)

In Study 1, participants completed four tasks<sup>5</sup>: a C-Test and a detailed self-assessment, each in both their L1 and their L2. Each of the two C-Tests comprised five short texts with a total number of 100 gaps determined by the schema proposed by Grotjahn (2010), and each correctly filled gap was awarded one point, so the maximum possible score in each language was 100. The self-assessments contained 43 5-point Likert-Scale items for the subdomains Listening (8 items), Reading (7 items), Speaking (17 items), and Writing (11 items). These items were constructed based on the ALTE Can-Do statements for levels C1 and C2 of the Common European Framework of Reference for Languages (CEFR, see Hulstijn, 2015: ch. 10). Responses were coded from 1 ("I cannot do this") to 5 ("I can do this without any difficulty"). Averages were created for the subscales as well as globally, with the maximum possible score being 5 and the minimum being 1.

Since the C-Test in the second language was different for the two populations (one being in English, the other in Dutch) the results were standardised for both groups by setting the lowest score in either population to 0 and the highest to 100. None of the tasks differed significantly across populations, although the Can-Do scales for the L1 approached significance, with the Dutch L2 speakers rating themselves somewhat higher than the English L1 speakers. The results are summarised in **Table 2**.

Tests of normality (K-S, bilingual data only) were significant at  $p < 0.01$  for all four variables, and visual inspection revealed all of them to be negatively skewed. The variables were therefore inverted and log-transformed, which resolved the normality issue for all except the L2 C-Test. That variable was root-transformed instead, resulting in a normal distribution. All variables were subsequently re-calculated to the same scale, so that in all cases the lowest score achieved was set to 0 and the highest to 1. The resulting standardised scores all correlated with the original scores above 0.95 (all  $p$ 's < 0.001). Lastly, we calculated an average score for both tasks in the L1 as well as in the L2 (both were normally distributed, with a lower bound of the significance at 0.2).

<sup>5</sup>All materials are available on <https://languageattrition.org/resources-for-researchers/experiment-materials/>

**TABLE 3 |** Response times (ms) Picture Naming Task in L1 and L2.

		PNT L1	PNT Dutch
Turks in the Netherlands (TRNL)	Mean	1,123	1,293
	Std	154	146
	Range	805–1454	1,000–1,601
Turkish controls (TRCG)	Mean	1,110	
	Std	136	
	Range	855–1442	
Moroccans in the Netherlands (MANL)	Mean	1,068	1,120
	Std	175	259
	Range	689–1400	747–1482
Moroccan controls	Mean	956	
	Std	155	
	Range	720–1284	
Dutch controls	Mean		895
	Std		121
	Range		620–1198

### Study 2: proficiency measures relating to basic language cognition (BLC)

Study 2 also used four instruments: Firstly, there were two Picture Naming Tasks (one in the L1, Turkish or Moroccan Arabic, and one in the L2, Dutch) in which participants were asked to say aloud the name of 78 objects which they saw as line drawings on a computer screen. The pictures were selected from the Snodgrass and Vanderwart (1980) dataset and controlled for cultural appropriateness, cognate status, item frequency and semantic and phonological relatedness between consecutive items (see Yilmaz, 2013 for details). Presentation was done through E-Prime 1.0 with a Serial Response Box and microphone to collect RTs, and all experiments were audio-recorded for later checking and verification of accuracy. Data were trimmed by eliminating all RTs below 250 ms as well as all items with inaccurate or missing responses. Outliers were defined as RTs higher than the mean plus two standard deviations, and these values were reduced down to the threshold for outliers. Based on these measures, the average RT for each participant in the L1 and in the L2 was calculated (see Table 3).

In this case, the L1 naming latencies of the two bilingual groups did not differ substantially from each other. Nevertheless, in order to ensure that language-specific differences would not impact on the results, we followed the same strategy for standardisation within the language groups as described above under Study 1 (based on the Full Dataset). L2 naming latencies were standardised only on the basis of the bilingual data and did not include the monolingual data.

The second set of variables was derived from a semi-structured interview conducted by a native speaker of the language in question (Turkish or Moroccan Arabic in Session 1, Dutch in Session 2) with no knowledge of the other language. Both interviews were autobiographical, informal, and focused on different aspects of the emigration experience. All interviews were transcribed and coded according to the guidelines set out in the CHILDES project for the CHAT format (MacWhinney, 2000).

The following variables were subsequently extracted from these transcriptions:

- **VOCD** (McKee et al., 2000). For the Turkish and Dutch data, VOCD was based on the lemmatised transcriptions by restricting the analysis to the %mor-tier. However, as there are no mor-grammars available for Moroccan Arabic, VOCD was calculated based on the main participant tier. In order to compensate for both the effect of language and the different methods of data extraction, VOCD in L1 was standardised separately for each speaker population. For Dutch VOCDs standardised scores were calculated based on all bilingual speakers (i.e., both Turkish and Moroccan participants).
- **Fluency**: for each participant, the number of filled pauses, repetitions and retractions (self-corrections) was counted and subsequently standardised to incidence per 1,000 tokens. Fluency variables were standardised across the groups for all speakers in the same way as described for the other variables above and averaged to one overall fluency measure per speaker in L1 and L2.

Tests of normality (K–S) showed no deviations from the normality assumption for any of the above variables. All were standardised to the same scale and direction (0 being the worst attained score and 1 being the best), and subsequently, one average measure for L1 and L2, respectively, was created by averaging the three subtasks.

## RESULTS

In order to assess to what extent the variables established above for personal background, use of L1 and L2, and attitudes may be used to predict the outcome variables we conducted two analyses for each dataset: a multivariate analysis of covariance (MANCOVA), which creates an overall model but also allows identifying the regression slopes associated with each predictor for each outcome variable (RQ1), and a discriminant function analysis (DA), which allows a non-linear assessment of the impact of the predictors (RQ2). Each analysis was conducted separately for HLC aspects of proficiency (Study 1) and BLC aspects of proficiency (Study 2) (RQ3).

## MANCOVAs

### Study 1: HLC Aspects of Proficiency

The first MANCOVA was conducted on the four standardised outcome variables from Study 1: C-Test L1, C-Test L2, Can-Do L1 and Can-Do L2. The six components identified by the PCA were entered as covariates. All components were entered together. Roy's Largest Root was significant for all components except *Contact* (see Table 4 for the full results). *Interactive Use* was significantly associated with both of the L2 measures, with a higher level of L1 use associated with a lower L2 C-Test score and L2 self-rating, but did not influence outcomes in the L1. The *Personal Background* component was associated with both self-assessments, a higher score on this component (reflecting higher age, longer length of residence and a lower educational

**TABLE 4 |** MANCOVA for HLC proficiency tasks in Study 1.

		L1							L2			
		GLM			C-Test L1		CanDo L1		C-Test L2		CanDo L2	
	Roy's largest root	<i>F</i> (4,89)	<i>p</i>	Partial $\eta^2$	<i>p</i>	Partial $\eta^2$	<i>p</i>	Partial $\eta^2$	<i>p</i>	Partial $\eta^2$	<i>p</i>	Partial $\eta^2$
Interactive use	0.446	9.925	<0.001	0.304***					<0.001	0.126***	<0.001	0.154***
Personal background	0.526	11.693	<0.001	0.333***			<0.001	0.076***			<0.001	0.099***
Perception	0.322	7.164	<0.001	0.248***	<0.001	0.140***	<0.01	0.096**				
Attitude	0.332	7.388	<0.001	0.250***	<0.01	0.108**	<0.001	0.177***			<0.05	0.048*
Professional use	0.293	6.509	<0.001	0.220***	<0.001	0.211***			<0.01	0.104**		
Contact	0.028	0.613	0.654	0.027								

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

level) leading to participants rating themselves better in L1 and poorer in L2 – without, however, this being reflected in the more objective proficiency scores yielded by the C-Test. *Perception* influenced both L1 measures, with participants who claimed their L1 had not changed and was at a high level at the present time achieving better scores on the C-Test and rating their proficiency more positively. *Attitude* had the most consistent effect: participants for whom the maintenance of their L1 and its transmission to the next generation was more important and who stated a preference for their home culture achieved better scores on the L1 C-Test and also rated themselves more positively in both languages. A higher level of use of the L1 for professional purposes and a lower level of use of the L2 was associated with higher C-Test scores in both languages. For the full models, effect sizes were medium to large, partial  $\eta^2$  values ranging from 0.22 to 0.33, while the individual associations were weak at best with partial  $\eta^2$ s between 0.05 and 0.2.

In order to assess to what extent including both languages in the model improves explanatory validity (RQ2), we repeated the MANCOVA twice. The first model included only the L1 measures as dependent variables, and the second only the L2 measures. Here, *Interactive Use* and *Personal Background* variables only became significant for the L2, while *Perception* was significant only for the L1. *Attitude* and *Professional Use* were significant predictors in both models, while *Contact* was not significant for either language. Except for the impact of professional L1 use on the L1, effect sizes were considerably lower than in the full model, ranging from 0.06 to 0.22 (see **Supplementary Table S3**).

## Study 2: BLC Aspects of Proficiency

The findings were much less revealing for the less controlled aspects of language use tested in Study 2. Here, the only component that yielded a significant result overall was *Interactive Use*, with a higher level of L1 use in informal contexts associated with slightly slower responses on the L1 PNT as well as a reduction in L1 disfluencies. Partial  $\eta^2$ s were around 0.11 (weak effect) for the individual measures and 0.27 (medium effect) for the overall model. A few other significant relationships were

observed (see **Table 5**), but in those cases the overall model was not significant. These findings confirm earlier studies showing that accounting for informal features of language attrition on the basis of background variables is highly problematic in models based on linear regression slopes (e.g., Schmid and Dusseldorp, 2010).

Repeating the MANCOVA for each language separately did not return a significant result for any of the predictors except *Attitude*, which was only significant in the L2 model ( $p < 0.05$ ; partial  $\eta^2 = 0.155$ ).

The findings from these two analyses are interesting in the light of previous investigations of L1 attrition in that they underscore that, while including results relating to performance in both languages can increase explanatory adequacy, analyses looking for linear regression slopes typically yield few results and have little predictive power. While the analyses of the HLC skills presented above are overall significant, the predictive power of the independent variables is limited and inconsistent, and effect sizes are weak. The situation is even worse with respect to BLC aspects of proficiency, where no coherent picture emerges at all.

## Discriminant Function Analysis (DA)

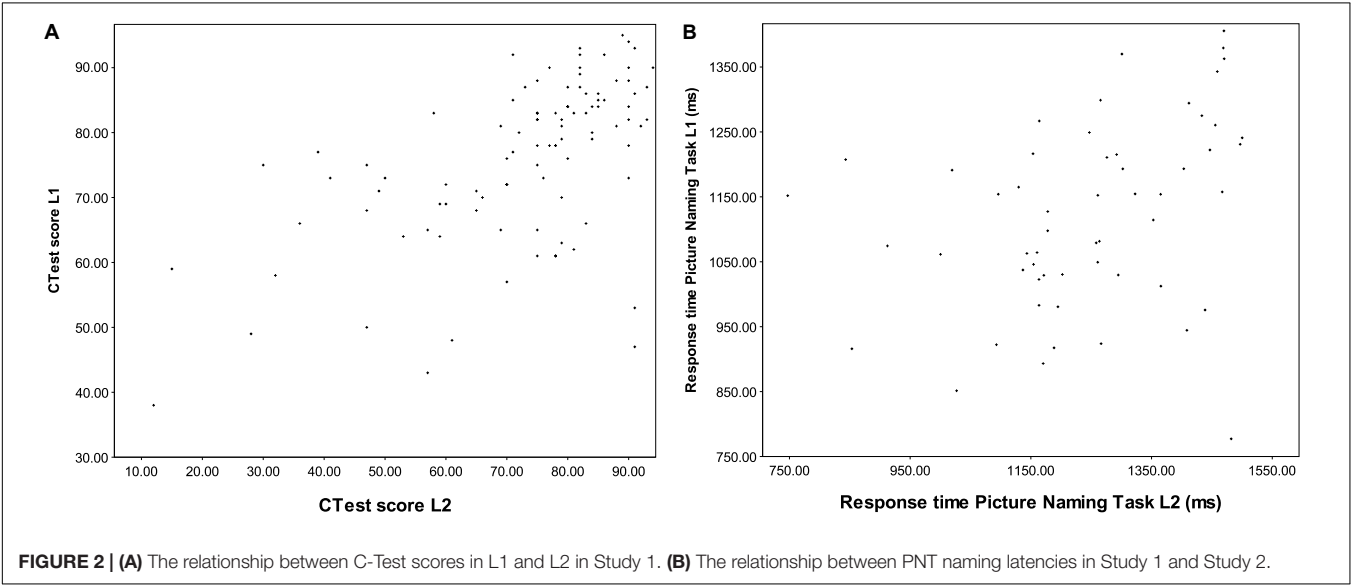
The very limited explanatory value of the two General Linear Model analyses described above points to a fundamental problem in research on bilingual development: the most common statistical analyses, such as regression or ANOVA, are only able to capture *linear* trends and correspondences in the data (i.e., correlation coefficients or regression slopes). In other words, any one predictor will only be revealed as significant if its impact on the outcome variable is the same for all or most of the participants. This can be seen, for example, in the fact that *Professional L1 Use* has the strongest impact on formal tasks such as the C-Test: it makes sense that individuals who engage with language as part of their job would develop enhanced awareness of style, orthography etc., facilitating these kinds of tasks. However, this relationship may not hold for all speakers (e.g., some speakers may retain excellent skills despite not ever using their L1 professionally) and may be far more complex for other types of background variables. For example, it is possible that some factors may interact with each other in non-linear ways



TABLE 5 | MANCOVA for BLC proficiency tasks in Study 2.

					L1						L2					
					RT		VOCD		Fluency		RT		VOCD		Fluency	
	Roy's largest root	F(6,43)	p	η²	p	η²	p	η²	p	η²	p	η²	p	η²	p	η²
Interactive use	0.369	2.648	<0.05	0.270*	<0.05	0.115*			<0.05	0.107*						
Personal background	0.103	0.736	0.623	0.093												
Perception	0.236	1.69	0.147	0.191	<0.05	0.113*									<0.05	0.113*
Attitude	0.216	1.552	0.185	0.178			0.051	0.077	<0.05	0.107*						
Professional use	0.095	0.678	0.668	0.086												
Contact	0.159	1.137	0.357	0.137			0.079	0.063								

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



(this has been suggested, for example, for length of residence and L1 use, de Bot et al., 1991; Schmid, 2011b).

A more serious problem for linear analyses is the fact that, as pointed out above (see section “Investigations of Development in Both Languages”) the balance between the L1 and the L2 is not the same for all members of the population: some speakers may preserve excellent proficiency in the L1 while also excelling in the L2, while for others, the development or maintenance of one language may come at the expense of the other, and others still may regress in their L1 without ever reaching very advanced levels of the L2. This variance cannot be captured by single-dimensional approaches, where the score obtained in one language is subtracted from that of the other, but it also eludes analysis by linear modelling. Consider the scores obtained on the C-Test in Study 1, and on the PNT in Study 2. For the former, there is a moderate positive correlation ( $r = 0.617$ ,  $p < 0.001$ ), indicating that participants who perform well in one language tend to perform well in the other (as pointed out above, this is probably due to the fact that general awareness of formal constraints facilitates this kind of task in any language), but there

are quite a few marked exceptions to this trend, as visualised in **Figure 2A**: for example, the two crosses toward the bottom right of the panel represent two participants who are among the lowest performers in the L1 but score within the top 15% in the L2. For the less controlled aspects of language use belonging to BLC aspects of proficiency, such as lexical access measured by the PNT and VOCD, on the other hand, no significant correlation between languages obtains, and no overall pattern can be detected from the scatterplot (**Figure 2B**).

In order to account for the four different types of bilingual balance identified in Treffers-Daller’s (2016, p. 261) ‘typology of language dominance based on language proficiency,’ analyses are therefore necessary that do not assume (negative or positive) linear relationships between development and proficiency in both languages. We propose a method here that proceeds from a *median split* of all participants in Study 1 and in Study 2. This was calculated on the basis of a single average score per language for the standardised variables C-Test and Can-Do Scales in Study 1 and PNT, VOCD and fluency in Study 2. Each participant was then categorised as having scored either above or below the

**TABLE 6 |** Subpopulations of bilinguals based on median split of proficiency scores.

	Study 1		Study 2	
	<i>n</i>	%	<i>n</i>	%
Good maintainer, good learner (performance above median in both languages)	36	34	18	28.6
Good maintainer, poor learner (performance above median in L1, below median in L2)	17	16	14	22.2
Poor maintainer, good learner (performance below median in L1, above median in L2)	17	16	12	19.0
Poor maintainer, poor learner (performance below median in both languages)	36	34	19	30.2

median in each language, yielding four groups (see **Table 6**). The division is visualised in **Figure 3A** (Study 1) and **Figure 3B** (Study 2).

The four-level categorical variables thus obtained for Study 1 and Study 2, respectively, were subsequently used as the grouping variable in two Discriminant Function Analyses (DA). DA attempts to find the best combination of predictors based on which as many cases as possible can be classified into the predetermined categories (Huberty and Olejnik, 2006). It is useful for the present investigation because (a) category membership is a nominal variable and thus does not imply any form of linear relationship or ranking and (b) the number of predictors is not limited based on the number of observations, as is the case, for example, in regression analyses and (M)ANCOVAs. The same 21 variables that were used for the Principal Component Analysis described above (see section

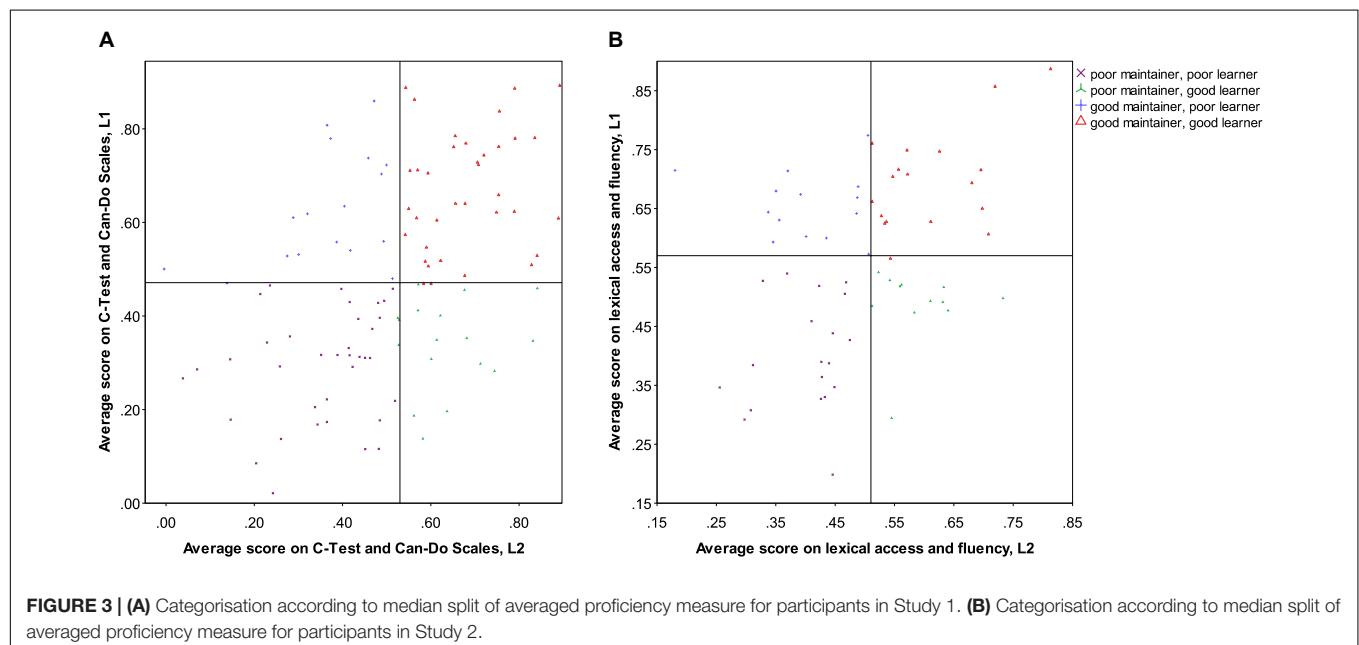
**TABLE 7 |** Discriminant Analysis Study 1, Functions at group centroids.

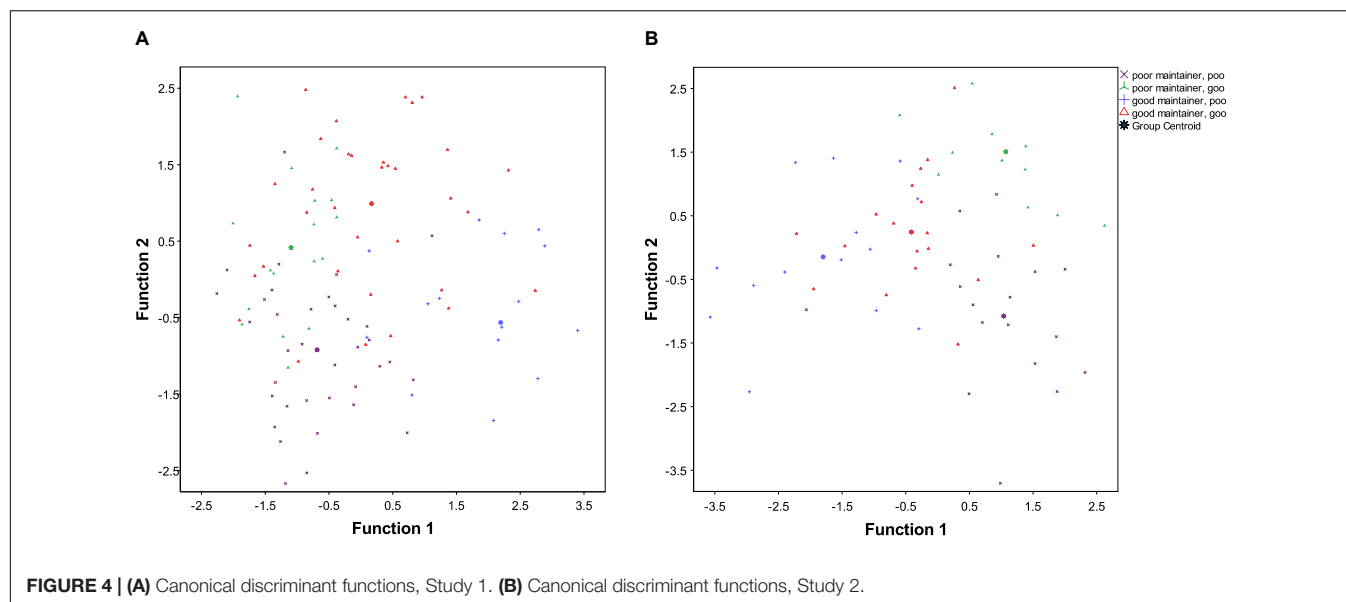
	Function 1	Function 2	Function 3
Poor maintainer, poor learner	−0.684	−0.922	0.277
Poor maintainer, good learner	−1.096	0.415	−0.976
Good maintainer, poor learner	2.195	−0.559	−0.362
Good maintainer, good learner	0.165	0.99	0.355

“Principal Component Analysis”) were entered as predictors into the DA.

For Study 1, three functions were identified which together significantly discriminated the four groups [Wilks'  $\lambda = 0.212$ ,  $\chi^2(63) = 143.297$ ,  $p < 0.001$ ]. The first function explained 54.6% of the variance (canonical  $R^2 = 0.54$ ), the second explained 33.7% (canonical  $R^2 = 0.42$ ) while the third explained only 11.7% (canonical  $R^2 = 0.20$ ). The cutoff point for factor loadings was set at 0.3 (the same threshold as used for the PCA).

**Function 1:** The factors loading on the first function mainly related to overall, interactive and informal use: the strongest factor here was the use of the L1 within the family, while the use of the L2 with friends and at work loaded negatively on this factor. The length of emigration was also a negatively loaded factor. An inspection of the group centroids (**Table 7**) suggests that this function was mainly associated with the maintenance of the L1: irrespective of their level of success in the L2, good maintainers tended to score positively on this function (that is, to have comparatively high levels of use of the L1 and low levels of use of the L2 in the contexts listed above, and short periods of residence), while poor maintainers scored negatively. This tendency is illustrated in **Figure 4A**, which also reveals it to be more pronounced for the poor





**FIGURE 4 | (A)** Canonical discriminant functions, Study 1. **(B)** Canonical discriminant functions, Study 2.

maintainers: while a number of the good maintainers/good learners score in the negative space of Function 1, none of the poor maintainers fall into the positive half of the chart, and only very few of the good maintainers/poor learners fall into the negative one. This suggests that good maintainers/good learners may possess high levels of aptitude, which allow them to attain high levels of proficiency in the L2 and overcome the negative impact of low levels of L1 use, high levels of L2 use and/or long periods of residence, retaining high levels of proficiency in the L1.

**Function 2:** The second function mainly differentiates good vs. poor learners (see **Figure 4A**), although there also seemed to be some impact on maintenance in that both among the good and among the poor learners the good maintainers scored more highly than the poor maintainers. The strongest factor in this function was educational level. Somewhat surprisingly, the use of the L1 for professional purposes also seemed to have a positive impact on L2 acquisition as well as maintenance. This suggests that using the first language professionally may help develop a higher sensitivity to some linguistic properties which are helpful in acquiring L2 proficiency, in particular as concerns the relatively controlled and monitored HLC skills measured in Study 1. Lastly, self-evaluated proficiency in the L2 also loaded on this factor.

**Function 3:** The last function was mainly associated with attitudinal factors. Variables loading onto this function were self-perceived changes to the L1 and current proficiency in the L1, the language and culture of preference and the proportion of friends with German as their L1. Interestingly, this function seemed to be associated with balanced vs. unbalanced bilingualism: those bilinguals who scored low or high in both their languages had a higher score on this function than those for whom one language was stronger (good maintainers/good learners scoring somewhat higher than poor maintainers/poor learners). The group of poor

maintainers and good learners had the lowest score on this function, followed by the good maintainers and poor learners. This tentatively suggests that a positive attitude toward the native language may support the development of the L2 toward its full capacity, matching that attained in the L1.

Based on these three functions, the DA was able to accurately classify 70.8% of original cases. In other words, 70.8% of all participants were assigned to the same of the four groups listed above by the DA and by the median split (see **Supplementary Table S4**).

The DA for Study 2 also identified three functions which together significantly discriminated the four groups [Wilks'  $\lambda = 0.143$ ,  $\chi^2(63) = 90.376$ ,  $p < 0.05$ ]. The first function explained 48.6% of the variance (canonical  $R^2 = 0.58$ ), the second explained 34.1% ((canonical  $R^2 = 0.49$ ) while the third explained 17.3% (canonical  $R^2 = 0.33$ ). Like in Study 1, Function 1 distinguished good and poor maintainers, while Function 2 distinguished good and poor learners (see **Table 8** and **Figure 4B**).

Function 1 comprised the language of preference, the language used with friends and the self-evaluated proficiency in the L2. As in Study 1, this function seemed mainly related to success in L1 maintenance, suggesting that people who preferred the L1, used the L2 less with friends and estimated their L2 proficiency lower were better maintainers (see **Figure 4B**).

Function 2 was again positively associated with level of education but also with self-assessed proficiency in the L1, and discriminated good and poor learners. Interestingly, good learners have a higher estimate of their own L1 proficiency than poor learners. Several other factors linked to the frequency of use of the L1 were positively associated with this function, but fell below the .3 threshold.

Function 3 comprised professional use of either language, with good maintainers/good learners reporting the highest levels

**TABLE 8 |** Discriminant Analysis Study 2, Functions at group centroids.

	Function 1	Function 2	Function 3
Poor maintainer, poor learner	−0.333	0.093	1.012
Poor maintainer, good learner	−1.692	0.299	−0.664
Good maintainer, poor learner	1.481	1.358	−0.342
Good maintainer, good learner	0.745	−1.385	−0.301

here and poor maintainers/poor learners the lowest, while the unbalanced groups had an intermediate position. This suggests that professional interactions – irrespective of the language in which they take place – may be beneficial for overall language development, possibly through the addition of a distinct domain for language use.

Together, these three functions accurately predicted group membership in 76.7% of original cases (see **Supplementary Table S5**).

## DISCUSSION

With the analyses presented above we have attempted to break new ground for the study of L1 attrition and language dominance. The knowledge gap we have addressed relates to the role of predictors in L1 attrition and the fact that, at the current state of knowledge, the empirical base for explanatory models of variability in L1 proficiency among immersed bilinguals is extremely weak. In other words, while we know that some individual speakers have attrited to a far higher degree than others, we do not know why. We therefore attempted to assess what circumstances in the environment of a particular speaker will facilitate the attrition vs. the maintenance of the L1. In order to do this, we adopted a novel approach. This proceeded from the assumption that explanatory models of language development, based on predictors comprising personal background factors as well as measures relating to exposure, use and attitudes, would be more powerful and more enlightening when both of the languages of the populations under investigation are taken into account.

In order to do this, we conducted two studies. The first one used linguistic measures related to Higher Language Cognition (mainly measuring participants' ability to manipulate language in ways that are not part of spoken daily interaction, through performance on a C-Test and self-ratings of language skills in a range of domains), while the second investigated the development of Basic Language Cognition, in particular in relationship to lexical access (Hulstijn, 2015). Both studies assessed these measures in both the participants' L1 (in Study 1, this was German, in Study 2 it was Turkish and Moroccan Arabic) and their L2 (Study 1: English and Dutch, Study 2: Dutch).

The first analysis attempted to identify linear relationships between the outcome measures on the one hand and the predictors on the other. It was demonstrated that including dependent variables relating to proficiency in both languages can considerably improve the explanatory validity of such models.

With respect to the formal tasks relating to HLC proficiency measured in Study 1, our results showed an impact of frequency of L1 use only for the L2 tasks, while aspects of personal background (such as age, education and length of residence) and introspective measures of proficiency and attitudes seemed to be reflected mainly in the self-ratings elicited by the Can-Do Scales in both languages. In the second analysis, findings were even more scarce and only suggested that higher levels of L1 use might have a facilitating – albeit very weak – effect on Reaction Times in lexical naming and fluency in informal speech in the L1.

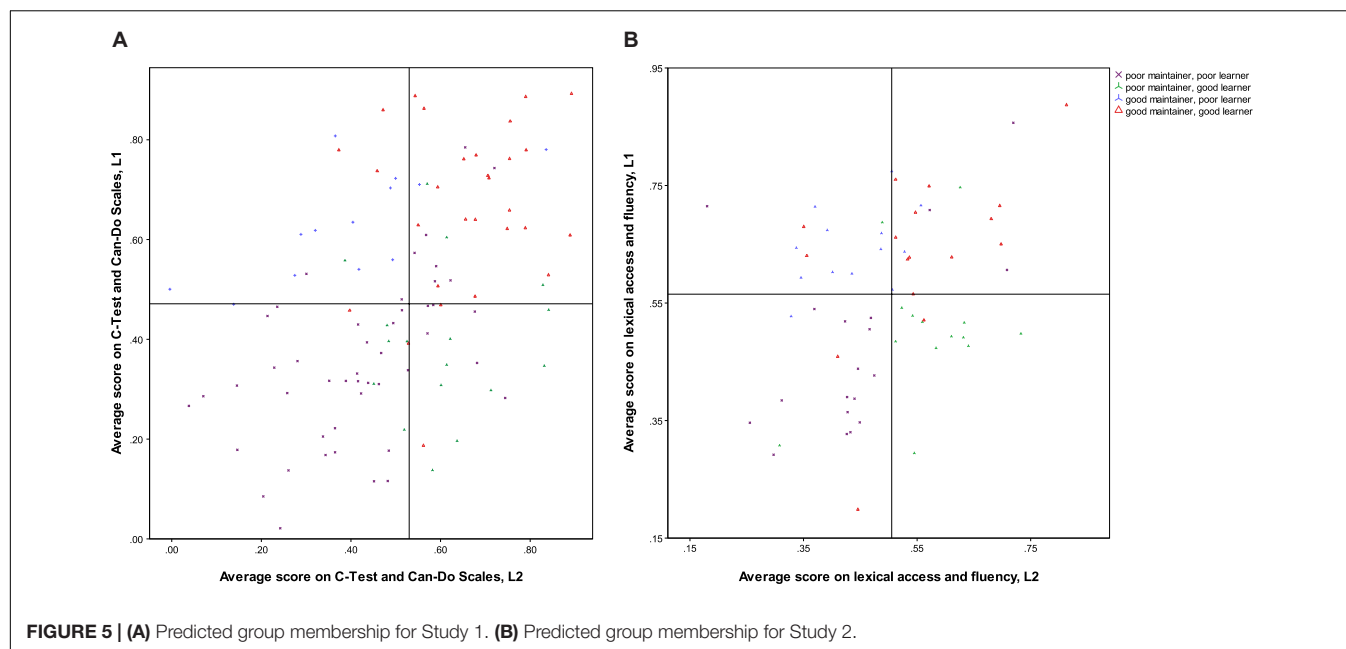
The somewhat disappointing results from this analysis are fully in line with previous work on L1 attrition: as was pointed out above, few studies have been able to identify any consistent impact, let alone any strong explanatory power, of predictors on actual measures of L1 proficiency and performance.

For our second analysis, we therefore adopted a different approach. Firstly, we combined the different measures of proficiency (two measures per language in Study 1, three in Study 2) into one compound measure. This was done following Opitz (2011, 2013) who showed that group differences which are masked in analyses based on single tasks may emerge when a compound measure is created. Secondly, we classified each participant into one of four quadrants of the proficiency space, based on whether they had performed above or below the median in each of their languages. This resulted in the creation of four distinct types of developers: good maintainers/good learners, good maintainers/poor learners, poor maintainers/good learners and poor maintainers/poor learners.

We fully acknowledge that this classification suffers from a number of problems that categorisation of interval data invariably entails: firstly, there is substantial loss of variance incurred by collapsing all of these different scores into just four categories. Secondly and relatedly, it results in the classification of those cases who are closest to the (arbitrary) threshold established for the cutoff into one group, even though they are far more similar to individuals on the other side of the threshold in another group than to many cases in their own category. This becomes evident from the visualisation of the categorisation in **Figures 3A,B**, above: the area in the middle of each chart contains participants whose scores in both languages are very close to each other, but who were assigned to different groups. We feel, however, that the benefits of capturing a relationship between the two languages that may go hand in hand for some participants but be orthogonal for others outweigh these drawbacks, but we would be delighted to learn of other analyses that are able to achieve this without resorting to categorisation of data.

Given the lack of previous insights into what factors may predict the development of a native language in immersed bilinguals, the insights gained from this classification can only be described as both unexpected and dramatic. Our hypothesis that treating language proficiency as a two-dimensional construct was confirmed by the Discriminant Analysis which was able, in both studies, to classify around three quarters of all participants accurately. Given the substantial differences between the two studies, both in terms of the population and of the linguistic skills analysed, it was particularly striking that, in both cases, the first – and hence most powerful – of the three functions





identified in the analysis related to L1 maintenance (irrespective of attained level of L2 proficiency) and comprised mainly measures related to informal language use: in both studies, those participants who had higher levels of language maintenance were the ones who used the L2 less with their friends and the L1 more with their family. This finding lends support to the often intuitively held view that more informal use of the L1 should be conducive to L1 maintenance – which, however, so far has lacked empirical substantiation. For example, a study on L1 attrition of lexical access and fluency measures very similar to the ones investigated in Study 2 here and using the same set of predictors (Schmid and Jarvis, 2014) finds no impact whatsoever of any factors linked to exposure. In a similar vein, a multivariate analysis of measures similar to the ones used in both studies here is presented by Schmid and Dusseldorp (2010). In the absence of the dimension presented by the measures in the L2, they conclude that “[l]anguage use in the more informal settings appears to have very limited protective function with respect to L1 attrition” (p. 152; see Schmid, forthcoming for a review of research attempting to link L1 use and L1 attrition).

Interestingly, the first function returned by the DAs, while successfully separating good and poor maintainers, seemed unrelated in both studies with success obtained in an L2, suggesting that the amount of informal use a participant makes of both her languages does not play a strong role when it comes to the development of either HLC or BLC skills in a second language. Here, it was found across both studies that the level of education as well as the level of self-perceived proficiency seemed to play a role. In Study 1, this function also affected L1 maintenance to some extent, suggesting that, when it comes to HLC, a higher level of education may be beneficial not only for L2 acquisition but also for L1 maintenance. Similarly, in this study, the use of

the L1 at work was important for both successful acquisition and successful maintenance. In Study 2, the effects of these factors were less pronounced, which is hardly surprising given that the BLC skills investigated in this study are probably much less amenable to educational levels. A rather puzzling finding is that the strongest contributing factor here was self-assessed L1 proficiency, with higher levels of proficiency being associated with better L2 skills – again, this may be related to the (unassessed) individual difference of language learning aptitude, which may have facilitated both L2 acquisition and L1 maintenance.

The last function was the only one for which there was no common pattern across the two studies. In Study 1, it seemed that a more positive attitude toward the native language and its maintenance would contribute to a more balanced pattern of language dominance, while in study 2 it seemed that using either the L1 or the L2 professionally may facilitate a higher level of proficiency in both languages.

It thus seems that the complex interaction of the predictors of both L1 attrition and L2 acquisition can be captured better by analyses which (a) plot out their results in a fully two-dimensional fashion and (b) do not rely on sweeping averages of the many necessary predictors, as we did through the Principal Component Analysis which yielded the independent variables used in the first set of analyses (MANCOVAs). As we pointed out above, the categorisation of data has a number of undesirable results, as it assigns cases which are very similar to each other to different groups. However, a closer look at the classifications yielded by the DA suggests that the negative impact of this may be less dramatic than one might have thought: **Figures 5A,B** depicts the median split division that was shown in **Figure 3** above. However, in this case, the markers showing the position of each individual do not represent their original group membership, but the group assigned to them by the DA.

In both studies, it is striking that, with few exceptions, most of the misclassified cases occur quite close to the median split lines. Recall that the DA does not have access to the actual scores of any of the individuals, only to the categorical group membership data. The fact that, even so, of the roughly 25% of misclassified individuals the vast majority are to be found among the more marginal cases underscores the potential value of such an analysis.

In order to gain more insight into the mechanisms of the development of language dominance, it may then be beneficial to adopt the approach suggested by Opitz, and scrutinise the more extreme cases of misclassification. We would like to illustrate this approach with the example of one individual participant in Study 2. This speaker attained the second highest score in the L1 and joint second highest in the L2, but the DA predicted her to be a poor maintainer and poor learner (represented by the purple cross toward the top right of the panel in **Figure 5B**). The participant in question is a Turkish woman who had come to the Netherlands aged 18 and, at the time of the data collection, had been living there for 27 years. While she used the L1 almost exclusively in her social life, she had in the early years of her emigration had contact with some Dutch women who had begun to teach her that language. She found that she very much enjoyed learning the language and, just before the time of data collection, had begun taking Dutch lessons for the first time (at the suggestion of her line manager). In the interview she talks about discovering aspects of the Dutch language and grammar that she had not previously been aware of, and what an enlightening and enjoyable experience this was for her. Furthermore, developing her Dutch skills also proved an empowering experience which changed her relationship to her overbearing and somewhat authoritative husband. It thus seems that, for this participant, a number of factors not measured in the present study, but probably relating to a high level of language aptitude and the experience of personal growth and self-fulfilment offered by the development of her linguistic skills, was enough to override the combination of the factors based on which the DA predicted low achievement in both languages for her.

As this and other cases in which the DA was unable to predict group membership show, the factors we included in our research design are not sufficient to paint a full picture of the circumstances under which both L1 maintenance and L2 acquisition may be more or less successful. Future studies should delve yet deeper into these questions and attempt to measure

personal characteristics, such as language aptitude, and other aspects of attitude and motivation.

What the study presented above shows very clearly, however, is that investigations of language dominance cannot afford to adopt a one-dimensional perspective, nor to rely on linear models of predictor-outcome relationships. We hope that these findings may inform future studies and also encourage investigations that are able to zoom in on more specific linguistic features than the relatively global and holistic ones we were able to measure here, in order to further inform our understanding of bilingual development and the forces that drive and shape it.

## AUTHOR CONTRIBUTIONS

MSS was responsible for data collection, classification and coding for Study 1, conducted all analyses reported here, and wrote the first draft of the article. GY was responsible for the data collection, classification and coding of the Turkish data in Study 2, and contributed to the revisions of the article.

## FUNDING

The research reported here was funded by the Dutch National Science Organisation NWO, NWO-grant 275-70-005 (Study 1), NWO-grant 360-70-250 (Study 2), and ESRC grant ES/M001776/1.

## ACKNOWLEDGMENTS

We are grateful to Farah Jamjam-van der Kooi for the collection of the Moroccan Arabic data, and to all of our research assistants for their help with the data processing and coding. We are also deeply indebted to Kees de Bot for his help and support throughout the projects reported here, and, indeed, throughout our entire careers.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01306/full#supplementary-material>

## REFERENCES

- Ammerlaan, T. (1996). 'You Get a Bit wobbly...'. *Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition*. Ph.D. dissertation, University of Nijmegen, Nijmegen.
- Bahrack, H. P., Hall, L. K., Goggin, J. P., Bahrack, L. E., and Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *J. Exp. Psychol. Gen.* 123, 264–283. doi: 10.1037/0096-3445.123.3.264
- Baladzhaeva, L. (2013). L1 attrition without L2 acquisition: a case of Russian-speaking immigrants in Israel. MA dissertation, University of Haifa, Haifa.
- Baus, C., Costa, A., and Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: a case for interactivity. *Lang. Cogn. Process.* 23, 866–888. doi: 10.1080/01690960801962372
- Bergmann, C., Nota, A., Sprenger, S. A., and Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *J. Phonet.* 58, 71–86. doi: 10.1016/j.wocn.2016.07.001
- Bergmann, C., Sprenger, S. A., and Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychol.* 161, 25–35. doi: 10.1016/j.actpsy.2015.07.015
- Bialystok, E. (2001). *Bilingualism in Development*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511605963

- Birdsong, D. (2016). "Dominance in bilingualism: foundations of measurement, with insights from the study of handedness," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 85–105. doi: 10.1017/CBO9781107375345.005
- Bongaerts, T., Van Summeren, C., Planken, B., and Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Stud. Second Lang. Acquisit.* 19, 447–465. doi: 10.1017/S0272263197004026
- Bylund, E., Abrahamsson, N., and Hyltenstam, K. (2009). The role of language aptitude in first language attrition: the case of pre-pubescent attriters. *Appl. Ling.* 31, 443–464. doi: 10.1093/applin/amp059
- Chamorro, G., Sorace, A., and Sturt, P. (2016). What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Biling. Lang. Cogn.* 19, 520–532. doi: 10.1017/S1366728915000152
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *J. Phonet.* 40, 249–268. doi: 10.1016/j.wocn.2011.10.007
- Cherciov, M. (2011). *Between Acquisition and Attrition: The Dynamics between Two Languages in Adult Migrants*. Ph.D. dissertation, University of Toronto, Toronto, ON.
- Cherciov, M. (2013). Investigating the impact of attitude on first language attrition and second language acquisition from a dynamic systems theory perspective. *Int. J. Biling.* 17, 716–733. doi: 10.1177/1367006912454622
- Cook, V. (2003). "The changing L1 in the L2 user's mind," in *Effects of the Second Language on the First*, ed. V. Cook (Clevedon: Multilingual Matters), 1–18.
- Cummins, J. (1991). "Interdependence of first-and second-language proficiency in bilingual children," in *Language Processing in Bilingual Children*, ed. E. Bialystok (Cambridge: Cambridge University Press), 70–89. doi: 10.1017/CBO9780511620652.006
- de Bot, K., Gommans, P., and Rossing, C. (1991). "L1 loss in an L2 environment: dutch immigrants in France," in *First Language Attrition*, eds H. W. Seliger and R. M. Vago (Cambridge: Cambridge University Press), 87–98. doi: 10.1017/CBO9780511620720.006
- de Houwer, A., and Bornstein, M. H. (2016). "Balance patterns in early bilingual acquisition: a longitudinal study of word comprehension and production," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 134–155. doi: 10.1017/CBO9781107375345.007
- de Jong, N. H., Groenhout, R., Schoonen, R., and Hulstijn, J. H. (2015). Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Appl. Psychol.* 36, 223–243. doi: 10.1017/S0142716413000210
- de Leeuw, E. (2008). *When your Native Language Sounds Foreign: A Phonetic Investigation into First Language Attrition*. Ph.D. dissertation, Queen Margaret University, Edinburgh.
- de Leeuw, E., Schmid, M. S., and Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Biling. Lang. Cogn.* 13, 33–40. doi: 10.1017/S1366728909990289
- Domínguez, L. (2013). *Understanding Interfaces: Second Language Acquisition and First Language Attrition of Spanish Subject Realization and Word Order Variation*. Amsterdam: John Benjamins. doi: 10.1075/lald.55
- Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Erlbaum. doi: 10.4324/9781410613349
- Dostert, S. (2009). *Multilingualism, L1 Attrition and the Concept of 'Native speaker'*. Ph.D. dissertation, Heinrich-Heine Universität, Düsseldorf.
- Flores, C. (2015). Losing a language in childhood: a longitudinal case study on language attrition. *J. Child Lang.* 42, 562–590. doi: 10.1017/S0305000914000233
- Gass, S. M., and Mackey, A. (2007). "Input, interaction, and output in second language acquisition," in *Theories in Second Language Acquisition: An Introduction*, eds B. VanPatten and J. Williams (New York, NY: Routledge), 175–199.
- Genevska-Hanke, D. (2017). "Intrapersonal variation in late L1 attrition and its implications for the competence/performance debate," in *Linguistik im Nordwesten: Beiträge zum 8. Nordwestdeutschen Linguistischen Kolloquium*, 13. – 14. November 2015, eds N. Levkovych and A. Urdze (Bochum: Brockmeyer).
- Gertken, L. M., Amengual, M., and Birdsong, D. (2014). "Assessing language dominance with the bilingual language profile," in *Measuring L2 Proficiency: Perspectives from SLA*, eds P. Leclercq, A. Edmonds, and H. Hilton (Bristol: Multilingual Matters), 208–225.
- Gregg, K. (2010). Shallow draughts: larsen-freeman and cameron on complexity. *Sec. Lang. Res.* 26, 549–560. doi: 10.1177/0267658310366582
- Grosjean, F. (1997). The bilingual individual. *Interpreting 2*, 163–187. doi: 10.1075/intp.2.1-2.07gro
- Grosjean, F. (2016). "The complementarity principle and its impact on processing, acquisition, and dominance," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 66–84. doi: 10.1017/CBO9781107375345.004
- Grotjahn, R. (2010). *The C-Test: Contributions from Current Research*. Frankfurt: Peter Lang.
- Hopp, H. (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing*. Ph.D. dissertation, University of Groningen, Groningen.
- Hopp, H., and Schmid, M. S. (2013). Perceived foreign accent in L1 attrition and L2 acquisition: the impact of age of acquisition and bilingualism. *Appl. Psycholinguist.* 34, 361–394. doi: 10.1017/S0142716411000737
- Huberty, C. J., and Olejnik, S. (2006). *Applied MANOVA and Discriminant Analysis*. Hoboken, NJ: Wiley. doi: 10.1002/047178947X
- Hulstijn, J. H. (2011). Language proficiency in native and non-native speakers: an agenda for research and suggestions for second-language assessment. *Lang. Assessm. Quart.* 8, 229–249. doi: 10.1080/15434303.2011.565844
- Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Biling. Lang. Cogn.* 15, 422–433. doi: 10.1017/S1366728911000678
- Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-Native Speakers. Theory and Research*. Amsterdam: John Benjamins.
- Karayayla, T. (2018). *Turkish as an Immigrant and Heritage Language in the UK: Effects of Exposure and Age at Onset of Bilingualism on Grammatical and Lexical Development of the First Language*. Ph.D. dissertation, University of Essex, Colchester.
- Kasparian, K. (2015). *The Case of the Non-Native-Like First-Language: Neurophysiological Investigations of First-Language Attrition and Second-Language Processing*. Ph.D. Dissertation, McGill University, Montreal.
- Kasparian, K., Vespignani, F., and Steinhauer, K. (2017). First language attrition induces changes in online morphosyntactic processing and re-analysis: an ERP study of number agreement in complex Italian sentences. *Cogn. Sci.* 41, 1760–1803. doi: 10.1111/cogs.12450
- Keijzer, M. (2007). *Last in First Out? An Investigation of the Regression Hypothesis in Dutch Emigrants in Anglophone Canada*. Ph.D. dissertation, Vrije Universiteit: Amsterdam.
- Kupisch, T., and van de Weijer, J. (2016). "The role of the childhood environment for language dominance: a study of adult simultaneous bilingual speakers of German and French," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 174–194. doi: 10.1017/CBO9781107375345.009
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analysing Talk*. Mahwah, NJ: Lawrence Erlbaum.
- Mägiste, E. (1992). Second language learning in elementary and high school students. *Eur. J. Cogn. Psychol.* 4, 355–365. doi: 10.1080/09541449208406193
- Major, R. C. (1992). Losing english as a first language. *Modern Lang. J.* 76, 190–208. doi: 10.1111/j.1540-4781.1992.tb01100.x
- McKee, G., Malvern, D., and Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary Linguist. Comput.* 15, 323–338. doi: 10.1093/lc/15.3.323
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *J. Phonet.* 32, 543–563. doi: 10.1016/j.wocn.2004.02.002
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the Age Factor*. Amsterdam: John Benjamins. doi: 10.1075/sibill.39
- Montrul, S. (2016). "Dominance and proficiency in second language acquisition and bilingualism," in *Language Dominance in Bilinguals:*

- Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35. doi: 10.1017/CBO9781107375345.002
- Opitz, C. (2010). L1 attrition and L2 acquisition: global language proficiency and language dominance in adult bilinguals. *Eurosla Yearbook* 10, 248–281. doi: 10.1075/eurosla.10.13opi
- Opitz, C. (2011). *L1 Attrition and L2 Acquisition in a Second Language Environment*. Ph.D. dissertation, Trinity College, Dublin.
- Opitz, C. (2013). A dynamic perspective on late bilinguals' linguistic development in an L2 environment. *Int. J. Biling.* 17, 701–715. doi: 10.1177/1367006912454621
- Opitz, C. (2016). Attrition profiles: what CDST can and can't tell us about attrition. *Paper Presented at the 3rd International Conference on Language Attrition*, Colchester.
- Opitz, C. (in press) "A complex dynamic systems perspective on personal background variables in L1 attrition," in *The Oxford Handbook of Language Attrition*, eds M.S. Schmid and B. Köpke (Oxford: Oxford University Press).
- Schmid, M. S. (2011a). *Language Attrition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511852046
- Schmid, M. S., and Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: the impact of extralinguistic factors. *Sec. Lang. Res.* 26, 125–160. doi: 10.1177/0267658309337641
- Schmid, M. S., and Fägersten, K. B. (2010). Disfluency markers in L1 attrition. *Lang. Learn.* 60, 753–791. doi: 10.1111/j.1467-9922.2010.00575.x
- Schmid, M. S., and Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Biling. Lang. Cogn.* 17, 729–748. doi: 10.1017/S1366728913000771
- Schmid, M. S., and Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Ling. Approach. Biling.* 7, 1–31. doi: 10.1075/lab.17058.sch
- Schmid, M.S. (2011b). "Contact x time: external factors and variability in L1 attrition," in *Modeling Bilingualism: From Structure to Chaos*, eds M.S. Schmid and W.M. Lowie (Amsterdam/Philadelphia: John Benjamins), 155–176. doi: 10.1075/sibil.43.11sch
- Schmid, M.S. (forthcoming). "Frequency and recency – the impact of amount of use and length of residence on L1 attrition," in *The Oxford Handbook of Language Attrition*, eds M. S. Schmid and B. Köpke (Oxford: Oxford University Press).
- Silva-Corvalan, C., and Treffers-Daller, J., (eds) (2016). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107375345
- Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174
- Treffers-Daller, J. (2016). "Language dominance: the construct, its measurement, and operationalization," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 235–265. doi: 10.1017/CBO9781107375345.012
- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "clozing" the gap. *Stud. Sec. Lang. Acquisit.* 33, 339–372. doi: 10.1017/S0272263111000015
- Tsimpli, I., Sorace, A., Heycock, C., and Filiaci, F. (2004). First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. *Int. J. Biling.* 8, 257–277. doi: 10.1177/13670069040080030601
- Varga, Z. (2012). *First Language Attrition and Maintenance among Hungarian Speakers in Denmark*. Aarhus: Aarhus University.
- Waas, M. (1996). *Language Attrition Downunder*. Frankfurt: Peter Lang.
- Yagmur, K. (1997). *First Language Attrition among Turkish Speakers in Sydney*. Tilburg: Tilburg University Press.
- Yilmaz, G. (2011). Complex embeddings in free speech production among late Turkish-Dutch bilinguals. *Lang. Interact. Acquisit.* 2, 251–275. doi: 10.1075/bct.48.04yil
- Yilmaz, G. (2013). *Bilingual Language Development Among the First Generation Turkish Immigrants in the Netherlands*. Ph.D. dissertation, University of Groningen, Groningen.
- Yilmaz, G., and Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *Mental Lexicon* 7, 249–274. doi: 10.1075/ml.7.3.01yil
- Yip, V., and Matthews, S. (2006). Assessing language dominance in bilingual acquisition: a case for mean length utterance differentials. *Lang. Assessment Quart. Int. J.* 3, 97–116. doi: 10.1207/s15434311laq0302\_2

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





# The Age Factor Revisited: Timing in Acquisition Interacts With Age of Onset in Bilingual Acquisition

Petra Schulz\* and Angela Grimm

*Institute for Psycholinguistics and Didactics of German, Goethe University Frankfurt, Frankfurt, Germany*

## OPEN ACCESS

### Edited by:

Cornelia Hamann,  
University of Oldenburg, Germany

### Reviewed by:

Natascha Müller,  
University of Wuppertal, Germany  
Cristina Maria Flores,  
University of Minho, Portugal

### \*Correspondence:

Petra Schulz  
P.Schulz@em.uni-frankfurt.de

### Specialty section:

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

**Received:** 05 May 2018

**Accepted:** 19 December 2018

**Published:** 14 January 2019

### Citation:

Schulz P and Grimm A (2019) The  
Age Factor Revisited: Timing  
in Acquisition Interacts With Age  
of Onset in Bilingual Acquisition.  
*Front. Psychol.* 9:2732.  
doi: 10.3389/fpsyg.2018.02732

In this paper, we investigate whether timing in monolingual acquisition interacts with age of onset and input effects in child bilingualism. Six different morpho-syntactic and semantic phenomena acquired early, late or very late are considered, with their timing in L1 acquisition varying between age 3 (subject-verb agreement) and after age 6 (case marking). Data from simultaneous bilingual children (2L1) whose mean age of onset to German was 3 months are compared with data from early second language learners of German (eL2) whose mean age of onset to German was 35 months as well as with data from monolingual children. To explore change over time, children were tested twice at the ages of 4;4 and 5;8 years. The main findings were that 2L1 children had an advantage over their eL2 peers in early acquired phenomena, which disappeared with time, whereas in late acquired phenomena 2L1 and eL2 children did not differ. Moreover, 2L1 children performed like monolingual children in early acquired phenomena but had a disadvantage in the late acquired phenomena with the amount of delay decreasing with time. We conclude that age of onset effects are modulated by effects of timing in monolingual acquisition. Contrary to expectation, input in terms of language dominance, measured as the dominant language used at home, did not affect simultaneous bilingual children's performance in any of the phenomena. We discuss the implications of our findings for the hypothesis that acquisition of late phenomena is determined by input alone and suggest an alternative concept: the learner's internal need for time to master a phenomenon, which is determined by its complexity and cross-linguistic robustness.

**Keywords:** language dominance, age of onset, timing in monolingual acquisition, language input, bilingualism, early second language acquisition, simultaneous bilingual acquisition, LiSe-DaZ

## INTRODUCTION

A growing body of research is devoted to child bilingual language learners (see Chondrogianni, 2018, for an overview). It complements the research on adult second language acquisition by examining age of onset effects among different types of child bilingual acquisition. The goal of our study is to contribute to the debate instigated by Tsimpli (2014) on whether age of onset effects can be modulated by effects of timing in monolingual acquisition. The concept "timing in acquisition" refers to the assumption that L1 development of the phenomena examined in bilingual children systematically modulates other factors such as age of onset and input. To address this issue, data were collected from simultaneous and early successive bilingual children acquiring

German as well as from monolingual German children. We asked whether simultaneous and early successive bilingual learners differ regarding early phenomena, with an advantage for the simultaneous bilinguals. We also asked whether (very) late phenomena result in similarly high or low performance, differentiating both groups of bilinguals from monolinguals, as predicted by Tsimpli's (2014) account. Furthermore, we addressed the issue of how timing in L1 acquisition interacts with differences in amount of language input by investigating the group of simultaneous bilingual children in more detail. This is a group that is often assumed to acquire target languages as their monolingual peers do and within a similar time frame as well (e.g., Genesee and Nicoladis, 2007; Paradis et al., 2011a).

The distinction between simultaneous and successive child bilingual acquisition hinges on the different outcomes postulated by acquisition theories. There is general agreement that acquisition of a second language after age seven qualitatively differs from first language acquisition, reaching the upper cut-off point for a critical or sensitive period for L2 acquisition (see Meisel, 2011, for an overview). At the lower end of the continuum, a consensus prevails that simultaneous acquisition of two languages from birth needs to be considered discretely as well, for this acquisition context falls within the sensitive or critical periods for language acquisition (e.g., Locke, 1997). Regarding the definition of successive childhood bilingualism there is less agreement. Some studies propose that age of onset to the second language occurs between the ages of 1 and 3 years (e.g., Unsworth, 2013a), while other studies suggest that age four constitutes an important cut-off point (Schwartz, 2004; Rothweiler, 2007; Meisel, 2009; Schulz and Tracy, 2011; Unsworth, 2016). Consequently, the upper age limit for what counts as simultaneous language acquisition varies considerably, ranging from birth to about 2 years (De Houwer, 2009). In addition to these theoretically inspired questions, age of onset issues are confounded with country-specific educational practices. This is because the age of onset to the second language often coincides with the age at which children tend to start daycare, where in general the country's majority language is spoken. In the present study, we use the term "bilingual acquisition" to refer to children who acquire two languages. Following our previous work (Schulz and Tracy, 2011; Grimm and Schulz, 2014a, 2016) the term "early second language acquisition" (henceforth also: eL2) is used to refer to children whose age of onset to the L2 is between the ages of 2;0 and 4;0 years; and the term "simultaneous bilingual language acquisition" (henceforth also: 2L1) is used to refer to children who are first exposed to the "other" language between birth and the age of 23 months. This way we capture the fact that children who are exposed to the second language after the age of 24 months have already developed substantial lexical and grammatical knowledge in their first language and cannot be considered "simultaneous learners" anymore.

In the current study we investigate language domains in morpho-syntax and semantics that comprise early and (very) late acquired phenomena in three different acquisition types. Although these acquisition types could all be called "early," they differ regarding age of onset and regarding input in a second

language. More specifically, we examine whether 2L1 children are like eL2 children. Both acquire German as one of two languages but differ in their age of onset. And we examine whether 2L1 children are like monolingual German-speaking children. Both have comparable ages of onset to German but have different amounts of input in the L2 German. To take into account the interaction of age of onset, timing in L1 acquisition and time of testing we adopted a longitudinal design in which data were collected in two test rounds at ages 4;4 and 5;8, about 16 months apart. To explore whether language dominance in the group of simultaneous bilingual children affects their performance, we determined subgroups of German-dominant, non-German-dominant and balanced simultaneous bilinguals, based on the dominant language used at home. Data on all phenomena were collected with the standardized test LiSe-DaZ (Schulz and Tracy, 2011). In short, our goal was to investigate the effects of age of onset (from birth, around age 3) and of timing in L1 acquisition (early, late, very late) in child bilingualism. Second, we explored whether language dominance, defined as the dominant language used at home, affects simultaneous bilingual children's performance.

## FACTORS INFLUENCING CHILD BILINGUAL ACQUISITION

Compared to monolingual children, bilingual children are subject to many more sources of variation in their language environment that may in principle influence their pace and path of acquisition. Factors such as general cognitive abilities and parental socioeconomic status play a role for some facets of children's language development. Additionally two sets of factors are especially relevant in the bilingual language acquisition context (Paradis and Jia, 2017; Chondrogianni, 2018): so-called age factors, related to the age of onset to the second language (see Section "Age of Onset Effects in Child Bilingualism"), and so-called input factors, related to input quantity and quality as well as to language dominance (see Section "Effects of Language Input and Language Dominance"). In some studies length of exposure to the second language has been classified as an input effect as well (e.g., Unsworth, 2016; Chondrogianni, 2018). However, length of exposure is related to age of onset as well as to the child's chronological age, often resulting in confounds between these factors. We return to this issue in the discussion. Recently, Tsimpli (2014) has proposed "timing in L1 development of the phenomena examined in bilingual children's performance" as an additional factor (see Section "Timing in L1 Acquisition"). Taking up the well-known observation that results for bilingual children may differ depending on the area of language being investigated, Tsimpli (2014) argues that early and late acquired phenomena result in different outcomes for the different types of bilinguals. According to her account, the linguistic factor "timing in L1 acquisition" needs to be taken into account to meaningfully address the role of age of onset and of language input.

Note that comparison of results in bilingualism studies is sometimes difficult, as studies have investigated different aspects of the acquisition process. Some have focused on the nature of

the acquisition path, asking whether the acquisition phases and their sequences are the same and whether the error patterns in each acquisition phase are caused by the same underlying acquisition principles (Meisel, 2009; Schulz, 2013; Unsworth, 2013a; Rothweiler et al., 2017; Schulz and Schwarze, 2017). Others have focused on the pace of acquisition, asking how fast specific acquisition stages are reached and at what age specific structures are mastered (e.g., Grimm and Schulz, 2012; Paradis and Jia, 2017). Still others have focused on the success of acquisition, asking whether successful acquisition, often referred to as native-like attainment, is possible (e.g., Kupisch and Rothman, 2018; see Schulz, 2012, for discussion of native-like attainment in general). Given the standardized nature of our data, our study focuses on questions of pace and success.

## Age of Onset Effects in Child Bilingualism

The presence of age effects in second language acquisition is uncontroversial. A notably robust finding is that child second language learners in general outperform adult second language learners (see the influential study by Johnson and Newport, 1989). This effect has been attributed to the existence of one or several critical periods (Locke, 1997; Meisel, 2009, 2013; see Birdsong, 2006, for an overview) as well as to other cognitive factors (e.g., Klein, 1996; Bialystok and Hakuta, 2010). Recently bilingualism research has started to address the role of age of onset within childhood bilingualism.

One line of research focuses on successive bilingual learners with different ages of onset. Many studies on child L2 acquisition (age of onset of 6 to 7 years) have found that child L2 learners perform much like adult learners and very differently from eL2 children. Evidence for parallels between child L2 and adult L2 has been reported for passive in German (Wegener, 1998) and for verb-second and subject-verb agreement in German (Haberzettl, 2005; Rothweiler, 2006; Chilla, 2008). However, in a study on passives in English, Rothman et al. (2016) found that child L2 learners outperformed children with an age of onset of 4 to 5 years which they attributed to so-called conceptual transfer from the L1 to the learners' L2. Many studies on eL2 acquisition (age of onset of 3 to 4 years) have found parallels between eL2 and monolingual children. For example, eL2 children were reported to perform in a similar fashion to monolingual children on subject-verb agreement and verb-second in German in a number of different studies, producing the same types of error patterns and showing a delay only regarding the age of mastery (Prévost, 2003; Tracy and Thoma, 2009; Tracy and Lemke, 2012; Grimm and Schulz, 2014b; Rothweiler et al., 2017; Schulz and Schwarze, 2017). eL2 children were also found to acquire interpretation of German wh-questions in a similar fashion to monolingual children, showing a delay of about 1 year (Schulz, 2013). In a study of subject-verb agreement and clitic placement in L2 French, however, Meisel (2008) found that eL2 children's errors were similar to those found in L2 adults.

A further line of childhood bilingualism research focuses on the comparison of simultaneous bilinguals (2L1) with other populations. Studies comparing 2L1 and eL2 children have

not found age of onset effects for the phenomena under consideration. In a study of Dutch neuter gender, Unsworth et al. (2014) found that 2L1 and eL2 children behaved alike regarding consistency of gender assignment and agreement. Given their selection of participants, the authors were able to consider length of exposure and age of onset separately and found that target-like gender marking was controlled by length of exposure rather than by age of onset. In a study of the comprehension of German wh-questions in 2L1 children and eL2 children Roesch and Chondrogianni (2016) did not find an effect of age on onset; differences between the groups were accounted for by length of exposure. Similarly, in a study of case and gender marking in noun phrases in German Hopp (2011) found strong correlations between length of exposure and eL2 children's performance, but no effect of age on onset.

Studies comparing 2L1 children to monolingual children have generally found that simultaneous bilinguals were not disadvantaged, acquiring the two languages in a similar fashion and at a similar pace as their monolingual peers (for an overview see Genesee and Nicoladis, 2007). Some studies reported specific acceleration effects, whereas other studies reported delays (see Hager and Müller, 2015, for a discussion of robust and non-robust domains in 2L1 children in comparison to monolingual children of the same language; see Müller, 2017, for a discussion of sources for acceleration and delay). Acceleration has been found for the area of morpho-syntax, with functional elements from the more developed language serving as a bootstrap for the acquisition of the functional elements of the other language (e.g., Gawlitzek-Maiwald and Tracy, 1996). In a similar vein, Kupisch (2006) argued for a "booster" effect in the development of determiners, caused by the bilingual children's ability to use their knowledge in one language when producing determiners in the other language. Delays have been mainly reported for the lexical domain, with a 2L1 child's vocabulary in either language being smaller than that of monolingual same-age peers and with less accurate performance on rapid lexical retrieval tasks (Bialystok, 2009). But delays have also been found for grammatical gender (see Gathercole and Thomas, 2009, for Welsh; Eichler et al., 2013, for German neuter gender) and for dative case/gender in German (Hager and Müller, 2015).

In short, whereas the existence of age of onset effects in second language acquisition is undisputed, many open questions remain, including which are the relevant cut-off points and whether the slogan "earlier is better" always holds. Also, effects of age of onsets do not occur in isolation; they are related to factors such as length of exposure and the age at which the learners were studied. The latter factor is important because it determines whether we can expect bilingual learners to have had enough time to catch up and acquire the specific phenomenon by the tested age. This in turn points to the factor "timing in L1 acquisition" to be considered next.

## Timing in L1 Acquisition

According to Tsimpli (2014), timing in L1 development of the phenomena examined in bilingual children's performance interacts with other factors such as age of onset and input.

Differentiating between early, late and very late acquired phenomena, Tsimplici proposes that this classification reflects the differing impact of narrow syntax. Early phenomena are core, parametric and narrowly syntactic, whereas late and very late phenomena involving syntax-external or language-external resources are not narrowly syntactic. Put differently, core grammatical properties are products of narrow syntax and exclude semantic effects. Macroparameters such as object verb directionality (OV/VO) and verb-second and their related microparametric options hence constitute the core component of language, which is acquired early. Phenomena that do not belong to the core are associated with components outside of narrow syntax. These late phenomena, acquired after age five, may involve semantics and pragmatics as well as non-verbal cognitive resources. More specifically, they may require knowledge at the semantic-syntactic interface and sensitivity to contextual information as with quantification and exhaustivity in wh-questions or they may require increased computational efforts as for instance in the comprehension of object-questions (Tsimplici, 2014: 293–294). Most importantly for the present study, it is argued “[...] that early phenomena can differentiate between simultaneous and (early) successive bilingualism with an advantage for the former group, while the other two reveal similarly (high or low) performance across bilingual groups, differentiating them from monolinguals” (Tsimplici, 2014: 283–284).

In the following we provide a classification of selected phenomena as early, late or very late, which have been studied in bilingual acquisition and which were tested in the present study (see Schulz, 2007; Schulz and Grimm, 2012, for an overview of the timing of acquisition in monolingual German). In line with Tsimplici (2014) we assume that earliness and lateness of the phenomena depend on whether additional resources or language submodules are involved, but we remain agnostic as to whether early phenomena have to belong to core-syntax. It may be that formal complexity plays a role, i.e., how much idiosyncrasy and irregularity is involved in a construction (see Culicover, 2014), which may or not may not align with the distinction between core and non-core (see also the contributions in Newmeyer and Preston, 2014).<sup>1</sup> Among the early phenomena, acquired before age 5, are object-verb directionality (OV/VO), verb-second and subject-verb agreement in German (Clahsen, 1986; Tracy, 1991) as well as acquisition of subordinate clauses (Bloom et al., 1989). These phenomena belong to the core expressing macroparameters and do not involve semantics. Furthermore, grammatical gender in Greek (Tsimplici, 2003) and telicity (Penner et al., 2003; see Schulz, 2018, for an overview) are acquired early. For gender in Greek it is argued that it shows consistent cues for gender values on nouns making it a grammatical gender language (Tsimplici, 2014: 298). For telicity, it could be argued that it involves mostly lexical knowledge and no resources at the level of sentential semantics. Among

the phenomena referred to as late, i.e., acquired around age 5, are passives (Armon-Lotem et al., 2015) and comprehension of relative clauses and wh-questions (Friedmann et al., 2009). According to Tsimplici (2014: 295) these phenomena require additional semantic or lexical information, with the possible exception of relativized minimality accounts of wh-movement (see the discussion in Friedmann et al., 2009). Finally, among the very late phenomena, acquired at age 6 and later are sentential negation (Wojtecka et al., 2011), exhaustivity in multiple wh-questions (Roeper et al., 2007; Schulz and Roeper, 2011; Schulz, 2015), grammatical gender in Dutch (Blom et al., 2008), and the case marking paradigm in German (Tracy, 1986; Eisenbeiss et al., 2005). Sentential negation and exhaustivity in wh-questions require semantic information and language-external resources. Grammatical gender in Dutch exhibits inconsistent cues for gender values on nouns and requires lexical knowledge (Tsimplici, 2014: 301). Similarly, the case marking paradigm in German exhibits intransparent cues for case marking on determiners, requiring lexical knowledge of the gender of the nouns and of the case suffixes within the tripartite gender system in German.

The few studies testing Tsimplici (2014) timing hypothesis confirm that timing differences result in different patterns for 2L1 and successive bilingual learners. Investigating the effects of age on onset and of input in grammatical gender in Greek (early) and Dutch (very late), Unsworth et al. (2014) found that amount of input was a predictive factor for the pattern attested in both Greek and in Dutch, whereas age of onset could explain the differences between 2L1 and successive bilinguals in Greek, but – as predicted by the timing hypothesis – not in Dutch. Likewise, the age of onset effects found by Meisel (2016) for gender in French, which is acquired early, are in line with the timing hypothesis. Furthermore, in a study with school-aged eL2 children acquiring English with a mean age of onset of 3 years, Chondrogianni and Marinis (2011) found effects of length of exposure rather than of age of onset for the acquisition of the late acquired structures wh-questions and passives. Last, 2L1 and eL2 children have been reported to differ in their comprehension of wh-questions in German, acquired late, with differences being accounted for by length of exposure rather than by age of onset effects (Roesch and Chondrogianni, 2016).

As mentioned at the beginning of Section “Factors Influencing Child Bilingual Acquisition,” differences and parallels may concern acquisition process and patterns, acquisition pace, and acquisition success. The focus of the present study is on pace, i.e., the question of how fast progress on the acquisition of a specific phenomenon is being made, and on success, i.e., the question of whether and at what age a specific phenomenon is acquired. Pace is typically measured quantitatively as the percentage correct in a given task across several time points, and success is typically measured via mastery (e.g., 90% correct) or via emergence of a phenomenon (e.g., first productive occurrence). Note that no matter which measure is chosen results are likely to vary to some degree depending on the specific task used. For example, case marking in German has been reported to be mastered late (Tracy, 1986; Schulz and Tracy, 2011; Schwarze, 2018) but also early

<sup>1</sup> The question of whether the distinction between core and non-core can account for the earliness and lateness of all phenomena across languages is beyond the scope of the present study.



(Roesch and Chondrogianni, 2016). Accordingly it is important to consider the specific task when reporting specific ages of mastery. Finally, when considering the effects of timing in L1 acquisition, age of testing is crucial because, as with the effects of age of onset, it necessarily determines whether we can expect monolingual children to have acquired this phenomenon by that age.

## Effects of Language Input and Language Dominance

Children who grow up bilingually receive less input in either language than their monolingual peers (Paradis and Genesee, 1996; Unsworth, 2013a). Nevertheless, simultaneous bilingual children have often been reported to acquire the two languages without delays compared to monolingual children (e.g., Paradis et al., 2011a). Accordingly, roughly half of the monolingual child's input seems to be sufficient for successful acquisition (Thordardottir, 2010). However, this observation leaves open the issue of how input quantity and quality influence acquisition patterns, pace, and success. Questions of quantity and quality of input have subsequently motivated much bilingualism research (Müller, 1990; De Houwer, 2009). The issue of input quantity is closely tied to questions of language dominance and investigations into which language used with a child is dominant and in which contexts a child receives her input. The issue of input quality is related to children's parental background, including whether parents (and siblings) are native speakers of the language in question and which socio-economic status or educational background parents have.

Assessment of input factors is difficult, however. They may change over time and they may be connected to child-related factors, such as the child's language use, language preference, language proficiency, and language output (e.g., Bohman et al., 2010; Schmeißer et al., 2015). Accordingly, the question of how to define reliable measures of quantity and quality of input has recently received increased attention (Paradis et al., 2011b; Unsworth, 2013b; Unsworth et al., 2014; Tuller, 2015; Roesch and Chondrogianni, 2016). Unsworth (2013b) argues for a calculation of cumulative length of exposure, in addition to current amount of exposure, in order to capture the sum of bilingual children's language exposure over time. Her results on the acquisition of gender in Dutch indicate that both cumulative and current amount of exposure predicted 2L1 children's performance. However, when 2L1 children were compared with monolinguals in terms of cumulative length of exposure, their scores were as high as (or higher than) the monolinguals.'

Independent of the specific measures used, differences in amount of input have often been shown to affect both bilingual children's language abilities and the rate at which they acquire various linguistic phenomena relative to monolinguals. For instance, rate of acquisition of vocabulary and morpho-syntax in English/French bilingual children seems to be affected by language input and use (Paradis et al., 2011b). Similarly, in studies of bilingual children a connection was found between amount of exposure and language development for vocabulary and morpho-syntax (Thordardottir, 2010; Hoff et al., 2012).

In a similar vein, a study on vocabulary acquisition showed that, provided sufficient exposure to the majority L2 language, children who switched dominance from the L1 to the L2 caught up to their monolingual peers at an even faster rate than simultaneous bilingual children (Hammer et al., 2008). However, some studies found amount of language input at home in the majority language to be unrelated to children's language performance (e.g., Chondrogianni and Marinis, 2011, for eL2 learners), one of the reasons being parents' low proficiency level in the majority language in which the children were tested (Chondrogianni and Marinis, 2011; Paradis, 2011). Similarly, in a study of placement of finite and non-finite verbs in 2L1 children in German, Schmeißer et al. (2015) found that language dominance and grammatical development were not positively related. In short, while it is undisputed that language input and language dominance play an important role for children's language outcomes, it is far from settled how different language domains are affected and whether input effects are the same in simultaneous bilinguals and in eL2 children.

## Research Questions

The aim of our study is to explore the three factors discussed above – age of onset, timing in L1 acquisition, and language input – by assessing the performance of 2L1 children and eL2 children as well as of monolingual children across early and (very) late phenomena. As for “age of onset” to German, we compare 2L1 children, who are similar to monolingual children in that they have roughly the same age of onset, to eL2 children. This way we are able to shed light on differences and parallels between two groups of child bilingual learners, which have been argued to constitute distinct acquisition types on theoretical grounds. As for “timing in L1 acquisition,” we compare early, late and very late acquired phenomena to see whether children's acquisition pace and success differs across differently timed phenomena. The early phenomena under investigation are subject-verb agreement and telicity, the late phenomena are complex sentences and wh-questions, and the very late phenomena are sentential negation and case marking. Timing is considered in relation to the age of testing, which may take place before or after this domain has been mastered by monolingual children. As for the third factor, “language input,” we study language dominance in the 2L1 group to find out whether children who are predominantly exposed to German at home benefit from a higher amount of input in terms of rate of acquisition.

Our first research question (Q1) addresses the effects of age of onset and timing in L1 acquisition and asks how the factors age of onset and timing in L1 acquisition affect the performance of simultaneous bilingual and early second language learning children. More specifically, we assessed the extent to which age of onset (from birth, around age 3) accounts for bilingual children's performance and whether timing in L1 acquisition (early, late, very late) interacts with age of onset. If bilingual children's performance is mainly attributed to effects of age of onset, two predictions can be made. First, even relatively small differences in age of onset between the 2L1 (AoO = 3 months in our sample) and the eL2 group (AoO = 35 months in our

sample) should result in differences between these two groups, with the 2L1 children performing better than the eL2 children. Exposure of the 2L1 group to the L2 German is longer than exposure of the eL2 group to the L2 German at any given point in time. We hence predict that the advantage of the 2L1 group in terms of an earlier age of onset holds independent of the specific phenomena investigated. Note that this does not imply that eL2 children always lag behind their 2L1 peers: if a specific phenomenon is assessed later in development, *after* the eL2 learners have mastered it, the advantage of the 2L1 over the eL2 learners would no longer be visible – at least in quantitative terms. The second prediction concerns the comparison of 2L1 and monolingual children. If age of onset is the crucial factor in determining children's performance, simultaneous bilinguals and monolinguals are expected to perform similarly, as length of exposure to German is by definition roughly the same in the two groups, and this pattern should be constant across development. If, however, timing in L1 acquisition interacts with age of onset effects, the patterns of behavior are expected to differ depending on whether the phenomenon in question is acquired early or late or very late. Here we apply Tsimpli (2014) proposal to the acquisition types 2L1 and eL2. Accordingly, 2L1 children are predicted to have an advantage over eL2 children for early acquired phenomena, whereas for late and very late acquired phenomena the two groups are predicted to perform similarly, and different from monolinguals. Again, it should be noted that these patterns may change with age: if testing of an early acquired phenomenon takes place later in development, *after* the eL2 learners have mastered it, the advantage of the 2L1 over the eL2 group will no longer be present, because both will have reached ceiling performance. Crucially, for early acquired phenomena the expected advantage for 2L1 over eL2 children is also predicted by the factor age of onset alone; for (very) late acquired phenomena, however, the factor timing leads to different predictions than the factor age of onset alone. Children's performance was assessed across six phenomena that varied with regard to their timing in L1 acquisition (early, late, very late). Quantitative measures via the mean scores achieved were used as well as qualitative measures, through assessing whether mastery in that domain was reached. To consider the role of the time of testing in relation to the factor timing in acquisition, data were collected across two test rounds.

Our second research question (Q2) asks whether language dominance affects simultaneous bilingual children's performance. We restricted the question to the group of 2L1 children, because they are likely to vary with regard to dominance, whereas eL2 learners of German at preschool age are most likely dominant in their L1. Under the assumption that input is especially crucial for the late acquired phenomena, simultaneous bilinguals who are predominantly exposed to German at home are expected to profit from this input and show an advantage over balanced or non-German-dominant simultaneous bilinguals especially in phenomena acquired at age 5 or later. For early acquired phenomena, which we hold to be less influenced by input effects, simultaneous bilingual children should not show differences according to their language dominance.

## MATERIALS AND METHODS

The data was collected in the course of two research projects, MILA (Grimm and Schulz, 2012) and *cammino* (Schulz et al., 2014). In both projects monolingual and/or bilingual language acquisition in child learners of German was examined in a combined cross-sectional and longitudinal design. The children were recruited between 2008 and 2013 in and around Frankfurt/Main, Germany. The current study reports the results of two test rounds, conducted at the ages of 4;4 years (test round 1) and 5;8 years (test round 2).

### Participants

The sample for test round 1 included 49 monolingual (MON) and 111 bilingual children, all of whom spoke German and one "other" language that differed across children.<sup>2</sup> Of these children, 37 monolingual and 103 bilingual children participated in test round 2. The bilingual children were further divided into a simultaneous bilingual (2L1) group and a group of early second language (eL2) learners of German according to their age of onset to German. The 2L1 learners had systematic contact to German and the "other" language before 24 months of age. The eL2 learners had an age of onset between 24 and 48 months of age. Background information was collected via a parental questionnaire and telephone interviews with the parents conducted in German or in the parent's L1. All children visited a German-speaking daycare center.

Subsequent to formal parental consent, children were included in the study if they scored at a standard value of 70 or higher in the non-verbal scales of the K-ABC (Kaufman et al., 2003) at 52.4 months of age ( $SD = 4.7$ ), if there was no assignment to speech-language intervention, and if according to their kindergarten teachers and their parents they showed age-appropriate language development.

**Table 1** provides the participant information. Across the three groups children's parents had a similar socio-economic background, with the exception of the fathers of the 2L1 children, who had a longer school education than the fathers of the eL2 children. Note that the majority of studies uses maternal educational background; we included information on father's educational background for the sake of completeness. Non-verbal IQ of the monolingual and the eL2 children did not differ, but the 2L1 group had a significantly higher non-verbal IQ than both the monolingual and the eL2 group.<sup>3</sup>

### Monolingual Children (MON)

The monolingual group consisted of 21 girls and 28 boys. All children were born in Germany. In 43/49 cases, children's parents

<sup>2</sup>The data at test round 1 were analyzed in Grimm and Schulz (2016) with a focus on language assessment. Due to the many different L1's attested in our sample, the role of the L1 is not considered further in the current study. In previous studies we did not find an effect of the L1 for eL2 children's performance in LiSe-DaZ scales (Schulz and Tracy, 2011; Schwarze, 2018; Wojtecka, 2018, unpublished). See also Tracy and Lemke (2012) and Tracy and Thoma (2009), who report that morpho-syntactic development was independent of the child's L1.

<sup>3</sup>We return to the factor non-verbal IQ in the results (see footnote 7).

**TABLE 1** | Mean values and standard deviations for background variables of participants.

	MON			2L1			eL2			
	N	M	SD	N	M	SD	N	M	SD	
Test round 1										
Child										
Age (months)	49	52.2	2.0	41	53.0	5.3	70	52.3	4.6	
AoO (months)	49	0	0	39	3.2	7.0	70	35.2	4.2	ABC
LoE (months)	49	52.2	2.0	39	49.6	9.1	70	16.5	6.4	BC
Non-verbal IQ	49	92.2	11.9	41	98.2	12.5	70	93.2	12.0	AC
Test round 2										
Age (months)	37	68.8	1.6	38	67.2	7.7	65	67.8	6.0	
AoO (months)	37	0	0	36	3.5	7.2	65	35.2	4.4	ABC
LoE (months)	37	68.8	1.6	36	63.8	10.7	65	31.9	7.7	ABC
Mother										
School education (years)	49	11.6	1.6	41	11.5	1.9	68	10.7	2.5	
LoR (years)	2	9	1.4	25	15.9	7.6	48	13.0	7.5	
Father										
School education (years)	48	11.2	2.5	39	11.7	1.7	67	10.5	3.0	C
LoR (years)	0	–	–	23	16.9	7.2	46	14.5	7.2	

AoO: Age of Onset; LoE: Length of exposure; LoR: Length of residence; A: significant difference between MON and 2L1, B: significant difference between MON and eL2, C: significant difference between 2L1 and eL2.

were also born in Germany, and in six families one parent was born in another country. In all 49 families German was the only home language and the only language the children acquired.

### Simultaneous-Bilingual Children (2L1)

The 2L1 group consisted of 18 girls and 23 boys. All children except for three were born in Germany. Out of the total of 41 families, in 18 families both parents were born in another country. In 12 out of these 18 families the parents were born in the same foreign country (most often Turkey, Afghanistan, Bosnia/Serbia), and in 6 cases the parents were born in different foreign countries. In 19/41 families one parent was born in Germany and the other parent was born in another country, and in one family both parents were born in Germany. For three families, information was lacking. Children acquired one of 17 different other languages, with Turkish and Russian being the most frequent, spoken by five children each. Age of onset was very homogeneous within the 2L1 group: for the majority of 2L1 children (32/41) age of onset was at birth. For two children, age of onset was after 0 and before 12 months; for seven children age of onset was between 12 and 23 months.<sup>4</sup>

### Early Second Language Learners (eL2)

The eL2 group consisted of 43 girls and 27 boys. All children except for one were born in Germany. In 49 out of the total of 70 families, both parents were born in another country. In 44/49 families the parents were born in the same foreign country (most frequently Turkey, Afghanistan, Bosnia/Serbia) and in 5/49 families the parents were born in different foreign countries.

In 8 families one parent was born in Germany, and in three families, both parents were born in Germany; for 10 families this information was lacking. Children acquired one of 28 different other languages, with Turkish being the most frequent, spoken by 15 children.

Age of onset was homogeneous within the eL2 group as well: 51/70 children had an age of onset between 34 and 40 months, corresponding to the age at which children in Germany typically enter daycare.

### Language Dominance of the Bilingual Participants

Given the organization of the two projects, a child's language dominance was determined during test round 1 based on a parental questionnaire targeting language use at home. This resulted in a three-way classification as German-dominant, balanced or non-German-dominant. More specifically, children's language dominance was calculated as a ratio of non-German and German use by the mother, the father and the child's siblings. The calculation was based on responses to the following questions: For the languages spoken in their home, each parent was asked *Welche Sprache(n) sprechen Sie mit Ihrem Kind?* "Which languages do you use when talking to your child?" In addition, we asked *Welche Sprache(n) sprechen die Geschwister, wenn sie miteinander sprechen?* "Which language(s) do the siblings use when talking to each other?" For each language that was named one point was awarded. This yielded a maximum score of 3 and a minimum of 0 in each of the two languages. Then for each bilingual child, the ratio of non-German/German was calculated with values ranging between 3 and 0.<sup>5</sup> If all family members exclusively used the "other" language when

<sup>4</sup> A re-analysis of the subgroup of the 34 2L1 children with an age of onset before 12 months did not change the results; this holds for the participant variables (chronological age, IQ) as well as for the test results.

<sup>5</sup> For single parents, the calculation of the ratio did not change; the maximum value in this case was 2.

**TABLE 2 |** Language dominance in the simultaneous bilingual children.

Language Dominance Ratio	2L1 children (N = 41)
<b>Non-German-dominant (n = 11)</b>	
3	3
2	6
1.5	2
<b>Balanced (n = 12)</b>	
1	12
<b>German-dominant (n = 18)</b>	
0.67	11
0.5	6
0.33	1

speaking with the child, the value was set at 3. The minimum value of 0 was reached if all family members exclusively used German when speaking with the child. A ratio of 1 indicates that family members used both the “other” language and German, when speaking with the child. Children were classified as “non-German-dominant” if the score was  $>1$ , as “German-dominant” if the score was  $<1$  and as “balanced” if the score was 1.

Calculation of the ratio for the eL2 group confirms the expectation that at age 4;4 most of the children (51/70) are predominantly exposed to their L1 at home, compared to only 13 balanced and 6 German-dominant eL2 learners. As a result, the factor “language use” was not considered further in the group of eL2 learners. The distribution in the 2L1 group is summarized in **Table 2**.

As can be inferred from **Table 2**, language dominance in terms of language use is evenly distributed in the group of 2L1 children: at age 4;4 11 children of the 2L1 group are dominant in the non-German L1, 11 children are balanced bilinguals, and 18 children are dominant in German. In the sub-group of 32 2L1 children with an age of onset before 12 months, the distribution was the same.<sup>6</sup> Note that this measure of language dominance in terms of use, assessed via parental questionnaire, differs from more fine-grained measures such as evaluating input quantity and quality via detailed questionnaires or direct assessment; we return to this issue in the discussion.

## Material

Children’s language performance was assessed with the standardized test LiSe-DaZ, administered in German (Schulz and Tracy, 2011), which offers separate norms for monolingual children and for eL2 children. The test was normed on 912 children (609 eL2 children and 303 monolingual children) across eight German states from diverse regions in Germany (see Schulz and Tracy, 2011: 86–87). The majority of the eL2 children spoke Turkish as their first language, followed by

Indo-Iranian languages such as Urdu, Kurdish, and Dari and by Russian (Schulz and Tracy, 2011: 88). Three subtests assess comprehension of central rule-based language phenomena: Verb meaning (semantics), Wh-questions (syntax, semantics), and Negation (syntax, semantics). Three subscales assess language production via an elicited production task in core areas of morpho-syntax: Complex sentences, Subject-verb agreement, and Case marking. These six scales were further considered for our analyses, for they provide maximum scores, which allow us to calculate mastery of acquisition. Five further sub-scales assess word classes including main verbs, modal and auxiliary verbs, prepositions, focus particles, and subjunctions (morpho-syntax, lexicon). These latter sub-scales were not considered further, as they assess the number of tokens produced, which indicates productivity but are not suitable as a measure of mastery.

The scales Verb meaning (12 test items) and Negation (12 test items) are based on a Truth-value-judgment task that elicits *Yes* or *No* responses. The scale Verb meaning assesses whether children are sensitive to the differences between telic and atelic verbs (see Penner et al., 2003). Contrasting true and false negatives, the scale Negation tests children’s knowledge of sentential negation (see Wojtecka et al., 2011). Using a question-after picture-design, the scale Wh-questions (10 test items) assesses knowledge of argument and adjunct wh-questions by asking children to respond to a wh-question with the correct part of a sentence (see Schulz, 2013). The production scale Complex sentences analyzes the most complex sentence types produced by the child (see Schulz and Schwarze, 2017; Wojtecka, 2018, unpublished) on a scale ranging from 1, indicating utterance of single word utterances only, to 4, indicating use of embedded sentences. A specific level was assigned if the child produced at least three utterances corresponding to that level. The scale Subject-verb-agreement assesses children’s knowledge that in German subject and verb have to agree in number and person (see Schulz and Schwarze, 2017). It is calculated in two steps. First, the number of all utterances containing a subject and verb (sum 1) and the number of all utterances containing a subject and verb with correct subject-verb-agreement (sum 2) are calculated. Second, a ratio is calculated by dividing sum 2 by sum 1, with the maximum score being 1.0. The sub-scale Case marking considers the total number of correctly realized case markings for accusative and dative in object positions and prepositional phrases. The maximum score is 9 (see Schwarze, 2018).

Following Tsimpli (2014) the factor “timing in L1 acquisition” is defined as the age at which specific phenomena are mastered in monolingual acquisition. Based on findings from previous research (see Section “Timing in L1 Acquisition”) the six phenomena studied here can be loosely classified as early, late or very late. However, as noted before age of mastery may vary to some degree depending on the specific task used. For the purposes of the current study we therefore calculated the age of mastery of all phenomena under investigation in a more precise fashion by considering the norming data for monolingual children from the LiSe-DaZ manual. Norming data are available for four age-groups: 3;00–3;11, 4;00–4;11, 5;00–5;11, 6;00–6;11

<sup>6</sup>Of these 15 (47%) were German-dominant, 11 (34%) balanced bilinguals and 6 (19%) non-German-dominant. Of the 7 children with an age of onset between 12 and 23 months, 3 (43%) children were German-dominant and 4 (57%) were non-German-dominant. Correlation between age of onset and language dominance was marginally significant,  $\chi^2(2) = 5.728$ ,  $p = 0.057$ ,  $\phi = 0.383$ .



(for the values, see Schulz and Tracy, 2011: 92, 95–97). The cut-off criterion for mastery was set at 90% (see Unsworth et al., 2014). Regarding the comprehension scales Verb meaning, Wh-questions, and Negation and for the production scale Case marking, we determined the age at which a mean of 90% of the test items were answered correctly. Age of mastery for the remaining two production scales was determined as follows. Regarding the scale Complex sentences, we calculated the age at which the mean score reached 90%. This score corresponds to a raw mean of 3.6 out of 4, with 4 expressing productivity of complex sentences in production. Regarding Subject-verb agreement, we calculated the age at which the correctness rate of subject-verb agreement (i.e., proportion of sentences with correct subject-verb agreement out of sentences with subject and verb) reached 90%. **Table 3** summarizes the relevant scales of LiSe-DaZ and the respective age of mastery in the norming sample.

In short, according to the norming sample Subject-verb agreement is mastered at age 3 (early), Verb meaning at age 4 (early), Complex sentences at age 5 (late), Wh-questions and Negation at age 6 (late), and case marking after age 6 (very late).

## Procedure and Statistical Analysis

The children were tested by trained student assistants in a quiet room in their kindergartens, and all test sessions were video-recorded. Later analysis was carried out by different research assistants trained on data analysis. As LiSe-DaZ does not provide norms for 2L1 learners, all analyses were based on raw scores. Descriptive statistics were calculated for the study sample characteristics and for the raw scores of the six sub-scales of LiSe-DaZ. One-way analyses of variance were used to determine whether the factor Group (monolingual, 2L1, eL2) differed in age, non-verbal IQ, and educational background of the parents. Separate Kruskal–Wallis-Tests (one test for each sub-scale) were used to compare the groups; non-parametric tests were employed because the data were not normally distributed. Significant main effects were followed by pairwise comparisons (Mann–Whitney-*U*-Tests) adjusted for significance by Bonferroni-corrections. Effect sizes were calculated manually as a dividend of the *z*-score (taken from the standard test statistic) and the square root of the overall number of participants:  $r = z/\sqrt{N}$  (Field, 2013: 248).

## RESULTS

### Effects of Age of Onset and Timing in Acquisition

First, mean raw values achieved across the two test rounds were computed for the three child groups (see **Table 4**).<sup>7</sup> Inspection of the mean values at test round 1 (age 4;4) reveals a uniform pattern of results for all six scales: the monolingual children achieved higher scores than the 2L1 children, which in turn achieved higher scores than the eL2 children. For test round 2 (age 5;8), we observed this pattern in four out of six scales: in the scales Complex sentences, Wh-questions, Negation, and Case marking the monolingual children achieved higher scores than the 2L1 children, which in turn scored higher than the eL2 children. In the scale Subject-verb meaning, the 2L1 group performed like the monolingual group and better than the eL2 group. In the scale Verb meaning the two bilingual groups achieved the same score, which was lower than the score achieved by the monolinguals.

As noted above (see Section “Research Questions”), differences between two learner groups may be absent because both have not acquired the phenomenon under investigation or for the simple reason that testing took place so late in development that both groups, e.g., eL2 learners and 2L1 learners, by the time of testing have reached ceiling performance. To distinguish these two scenarios, we coded for each scale whether the child groups achieved mastery in that domain (i.e., reaching the raw mean value corresponding to the 90% criterion in monolingual acquisition, see Section “Timing in L1 Acquisition”). **Table 4**

<sup>7</sup>Recall that there were significant differences between 2L1 and eL2 children with regard to paternal school education and between MON and 2L1 children as well as between 2L1 and eL2 children regarding non-verbal IQ (see **Table 1**). To explore the potential effect of these differences for the group results, additional statistical analyses were employed. Since differences were restricted to some of the groups and the data was not treated as a binary variable, multiple linear regressions comparing MON and 2L1 children and 2L1 and eL2 children were performed. There was no effect of paternal school education for the 2L1 and eL2 children. A significant effect of non-verbal IQ was found in only 7 out of 24 cases: in three cases for the MON and 2L1 children (Subject-verb agreement at T1, Verb meaning and Negation at T2) and in four cases for the 2L1 and eL2 children (Verb meaning at T1 and T2, Wh-questions and Case marking at T1). These results suggest that paternal education and non-verbal IQ did not systematically affect children's performance.

**TABLE 3 |** Scales of LiSe-DaZ, maximum score, size of norming sample, mean raw values (and standard deviations), mean percentage correct, and age of mastery in monolingual acquisition according to the norming sample.

Scale	Maximum score	Size of norming sample	Mean raw value (SD)	Mean Percentage correct	Age of mastery in monolingual acquisition
<b>Comprehension</b>					
Verb meaning	12	78	10.86 (1.88)	90.50%~	4
Wh-questions	10	63	9.57 (0.64)	95.70%~	6
Negation	12	63	10.89 (1.62)	90.75%~	6
<b>Production</b>					
Case marking	9	63	6.21 (1.76)	69.00%~	>6*
Complex sentences	4	73	3.8	95.00%#	5
Subject-verb agreement	1	84	0.95 (0.08)	95.00%+	3

~Percentage of test items solved correctly. \*The oldest group of six-year-olds did not reach the 90% criterion. #Percentage of complex sentences out of all sentences.

+Percentage of sentences with correct subject-verb agreement.

illustrates the results, with shaded cells marking those scales in which the respective child group reached mastery. Across both test rounds, the data for the monolingual group are in line with the ages derived from the norming sample. At T1 the 2L1 group reached mastery in Subject-verb agreement and Verb meaning, and at T2 also in Wh-questions. At T1, the eL2 group reached mastery in none of the six scales, and at T2 only in the scale Verb meaning. **Table 5** depicts the results of the inferential statistics.

Significant main effects were found in test round 1 and in test round 2 for all scales. Turning first to the comparison between 2L1 and eL2 children, pairwise comparisons show that at age 4;4

the 2L1 group performed better than the eL2 group in four out of six scales (Subject-verb agreement, Verb meaning, Complex sentences, Wh-questions), with effect sizes ranging from weak to moderate, and like the eL2 group in the scales Negation and Case marking. Importantly for our argumentation, this parallel between 2L1 and eL2 children cannot be attributed to the eL2 group having already acquired Negation and Case marking by age 4;4, because neither reached the score for mastery in these phenomena. At age 5;8 the pattern is different: the 2L1 group behaved like the eL2 group across all six scales, with all effect sizes being weak. As inspection of mastery in **Table 4** shows, this

**TABLE 4 |** Mean raw values, standard deviations and mastery for MON, 2L1, and eL2 children in the scales of Lise-DaZ ordered by age of mastery in monolingual acquisition.

Scale	MON			2L1			eL2		
	N	M	SD	N	M	SD	N	M	SD
<b>Test round 1 (age 4;4)</b>									
Subject-verb agreement (max 1)	49	0.97*	0.04	40	0.95	0.07	58	0.87	0.21
Verb meaning (max 12)	49	11.3	1.2	41	11.0	1.2	70	9.7	2.1
Complex sentences (max 4)	49	3.8	0.39	41	3.5	0.57	65	2.7	0.16
Wh-questions (max 10)	49	8.6	1.5	40	6.4	2.5	68	4.8	2.5
Negation (max 12)	49	10.1	1.8	40	8.0	2.0	66	6.8	2.3
Case marking (max 9)	49	4.2	2.3	41	2.1	1.8	68	1.3	1.6
<b>Test round 1 (age 5;8)</b>									
Subject-verb agreement (max 1)	36	0.99	0.01	36	0.99	0.03	63	0.92	0.42
Verb meaning (max 12)	37	11.9	0.4	36	11.4	0.9	64	11.4	0.8
Complex sentences (max 4)	36	4.0	0.0	36	3.8	0.4	64	3.6	0.7
Wh-questions (max 10)	37	9.2 <sup>#</sup>	0.9	36	8.7	1.6	63	8.1	1.7
Negation (max 12)	37	10.9	1.4	36	9.6	2.1	63	9.4	1.9
Case marking (max 9)	36	5.8	1.9	36	3.9	1.9	64	3.1	2.0

\*Shaded cells indicate mastery in this scale. <sup>#</sup>Mastery is almost reached, the mean value for mastery is 9.57.

**TABLE 5 |** Statistical outcome at T1 and T2 (main effect, pairwise comparisons and effect sizes) for the six scales of LiSe-DaZ.

	Main effect			Pairwise comparisons					
	Acquisition type			2L1 - MON		2L1 - eL2		eL2 - MON	
	N	H(2)	p	p	r	p	r	p	r
<b>Test round 1</b>									
Subject-verb agreement	147	19.69	<0.001	0.869	0.11	0.010	0.28	<0.001	0.39
Verb meaning	160	26.49	<0.001	0.476	0.15	0.004	0.30	<0.001	0.46
Complex sentences	159	55.42	<0.001	0.014	0.29	<0.001	0.36	<0.001	0.68
Wh-questions	157	60.98	<0.001	<0.001	0.45	0.015	0.27	<0.001	0.72
Negation	155	49.89	<0.001	<0.001	0.45	0.102	0.21	<0.001	0.65
Case marking	158	44.72	<0.001	<0.001	0.42	0.112	0.2	<0.001	0.62
<b>Test round 2</b>									
Subject-verb agreement	135	7.64	0.022	0.367	0.18	0.922	0.10	0.017	0.28
Verb meaning	137	14.04	0.001	0.061	0.27	0.819	0.11	0.001	0.37
Complex sentences	136	19.91	<0.001	0.164	0.23	0.076	0.22	<0.001	0.44
Wh-questions	136	15.70	<0.001	0.471	0.17	0.075	0.23	<0.001	0.39
Negation	136	17.75	<0.001	0.006	0.36	1.00	0.06	<0.001	0.41
Case marking	136	30.39	<0.001	0.003	0.39	0.242	0.17	<0.001	0.55

Significant results are given in boldface.

parallel between 2L1 and eL2 children can be partially attributed to the eL2 learners having caught up to their simultaneous bilingual peers, who have mastered Subject-verb agreement, Verb meaning, and Complex sentences by age 5;8. However, this is not true for the scales Wh-questions, Negation and Case marking, which neither the 2L1 nor the eL2 group has mastered at that age. As expected, the eL2 learners performed significantly worse than the monolingual children in all six scales across both test rounds, with effect sizes ranging from moderate to large at test round 1 and from weak to large at test round 2.

Turning to the comparison between simultaneous bilingual and monolingual children, at age 4;4, the 2L1 group performed like the monolingual group in only two scales (Subject-verb agreement and Verb meaning), whereas in four out of six scales (Complex sentences, Wh-questions, Negation, Case marking) the scores of the 2L1 group were significantly lower than those of

the monolingual group, with effect sizes ranging between weak and moderate. By age 5;8 the pattern has changed: 2L1 and monolingual children did not differ in four scales (Subject-verb agreement, Verb meaning, Complex sentences, Wh-questions), but in the two scales Negation and Case marking the 2L1 children still performed significantly lower than the monolingual children, with moderate effect sizes. In summary, age of onset alone cannot explain the unique profile exhibited by the changes in the 2L1 group from test round 1 to test round 2.

## Results for Language Dominance

Addressing the second research question of whether language dominance affects the performance of simultaneous bilingual children, we took the three-way classification as German-dominant, balanced and non-German-dominant (see **Table 2**) as a starting point. Effects of language dominance are most

**TABLE 6 |** Mean raw values, standard deviations for German-dominant and non-German-dominant 2L1 children in the scales of LiSe-DaZ.

	German-dominant			Non-German-dominant		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
<b>Test round 1</b>						
Subject-verb agreement (max 1.0)	18	0.95	0.07	11	0.96	0.07
Verb meaning (max 12)	18	11.3	0.97	11	10.6	1.4
Complex sentences (max 4)	18	3.4	0.51	11	3.5	0.52
Wh-questions (max 10)	18	6.2	2.48	11	5.9	2.6
Negation (max 12)	18	7.3	1.8	11	9.2	1.6
Case marking (max 9)	18	2.1	1.7	11	1.4	1.4
<b>Test round 2</b>						
Subject-verb agreement (max 1.0)	16	0.98	0.03	9	9.9	0.02
Verb meaning (max 12)	16	11.5	1.1	9	11.2	0.97
Complex sentences (max 4)	16	3.8	0.45	9	3.9	0.33
Wh-questions (max 10)	16	8.6	1.1	9	8.9	2.1
Negation (max 12)	16	9.6	1.8	9	9.9	2.6
Case marking (max 9)	16	3.9	1.4	9	3.7	2.2

**TABLE 7 |** Statistical outcome for German-dominant vs. non-German-dominant simultaneous bilingual children at test round 1 and test round 2 (Mann-Whitney-*U*-Test) for the scales of LiSe-DaZ.

Scale	<i>N</i>	<i>U</i>	<i>z</i>	<i>p</i>	<i>r</i>
<b>Test round 1</b>					
Subject-verb agreement (max 1.0)	29	87.5	− 0.567	0.611	0.11
Verb meaning (max 12)	29	68.0	− 1.49	0.173	0.28
Complex sentences (max 4)	29	98.0	− 0.052	0.982	0.01
Wh-questions (max 10)	29	95.0	− 0.182	0.877	0.03
Negation (max 12)	29	<b>41.0</b>	− 2.66	<b>0.008</b>	<b>0.49</b>
Case marking (max 9)	29	77.0	0.310	0.340	0.06
<b>Test round 2</b>					
Subject-verb agreement (max 1.0)	25	65.0	− 0.501	0.718	0.10
Verb meaning (max 12)	25	58.0	− 1.00	0.452	0.20
Complex sentences (max 4)	25	62.0	− 0.816	0.598	0.16
Wh-questions (max 10)	25	45.5	− 1.56	0.136	0.31
Negation (max 12)	25	57.5	− 0.834	0.419	0.17
Case marking (max 9)	25	66.0	− 0.346	0.760	0.07

Significant results are given in boldface.

likely to be observed in German-dominant vs. non-German-dominant children, whereas outcome for balanced bilinguals is less clear. Accordingly, we first compared the two extreme groups (German-dominant vs. non-German-dominant) using the Mann–Whitney-*U*-Test in order to detect an influence of language dominance. The results for the six scales across the two test rounds are summarized in **Table 6**; the results of the inferential statistics are given in **Table 7**.

As **Table 7** illustrates, at age 4;4 German-dominant and non-German-dominant simultaneous bilinguals differed only in the scale Negation, with the German-dominant children performing worse than the non-German-dominant children. In the other five scales, there was no significant effect for the factor group. At age 5;8, German-dominant and non-German-dominant simultaneous bilinguals did not differ in any of the six scales.<sup>8</sup> A comparison of all three sub-groups of simultaneous bilinguals confirmed this result: at age 4;4 significant main effects were found only for the scale Negation (Kruskal–Wallis-test,  $H(2) = 6.329$ ,  $p = 0.042$ ). *Post hoc* comparisons revealed that this effect is due to the difference between German-dominant children and non-German-dominant children (Mann–Whitney-*U*-Test, Bonferroni-adjusted,  $p = 0.039$ ,  $r = 0.39$ ). At age 5;8, no significant group differences were found. In summary, language dominance, measured via the languages spoken at home, did not result in differences between simultaneous bilingual children acquiring German.

<sup>8</sup>Unlike in other studies, in our sample of 2L1 children, German-dominant and non-German-dominant children did not differ regarding the educational background of parents (mothers:  $t(27) = -0.874$ ;  $p = 0.390$ ; fathers:  $t(26) = -0.470$ ;  $p = 0.642$ ).

## Summary of Main Results

The current study addressed two research questions. Research question (Q2) asked whether language dominance, measured via the languages spoken at home, affects the performance of simultaneous bilingual children. It was answered negatively, as German-dominant and non-German-dominant simultaneous bilinguals were found to not differ in any of the six scales of LiSe-DaZ at age 5;8 and at age 4;4 only differed in the scale Negation (and in the unexpected direction).

Research question (Q1) assessed the extent to which age of onset (from birth, around age 3) accounts for bilingual children's performance and whether timing in L1 acquisition (early, late, very late) interacts with age of onset. For ease of comparison, the results are summarized in **Table 8** for test round 1 (age 4;4 years) and in **Table 9** for test round 2 (age 5;8). The symbols in **Tables 8** and **9** indicate whether the groups differ statistically (=) and, if so, which group performed significantly better than the other (< or >) (taken from **Table 5**). The symbols do not express level of performance, i.e., two groups may not differ because they both exhibited ceiling performance or because they both have not yet mastered the domain targeted by this scale. To distinguish these two scenarios, we indicate for each group and scale whether mastery has been reached (taken from **Table 4**).

As can be inferred from **Table 8**, at age 4;4 the 2L1 children perform like their same-aged monolingual peers in Subject-verb agreement and Verb meaning, but worse than the monolinguals in the Complex sentences, Wh-questions, Negation and Case marking. Regarding mastery, by age 4;4 2L1 children, just like their monolingual peers, master Subject-verb agreement and Verb meaning, but different from their monolingual peers have not mastered Complex sentences. Both

**TABLE 8** | Summary of the pairwise comparisons and group mastery by language domain (test round 1).

Timing in monolingual acquisition	Age of mastery in monolingual acquisition	Language domain	2L1–MON	2L1–eL2	eL2–MON
Early	3	Subject-verb agreement	2L1 = MON	2L1 > eL2	eL2 < MON
	4	Verb meaning	2L1 = MON	2L1 > eL2	eL2 < MON
Late	5	Complex sentences	2L1 < MON	2L1 > eL2	eL2 < MON
	6	Wh-questions	2L1 < MON	2L1 > eL2	eL2 < MON
	6	Negation	2L1 < MON	2L1 = eL2	eL2 < MON
Very late	> 6	Case marking	2L1 < MON	2L1 = eL2	eL2 < MON

>: sig better than, <: sig worse than, =: n.s., shading indicates that this group has reached mastery in this scale at this test round.

**TABLE 9** | Summary of the pairwise comparisons and group mastery by language domain (test round 2).

Timing in monolingual acquisition	Age of mastery in monolingual acquisition	Language domain	2L1–MON	2L1–eL2	eL2–MON
Early	3	Subject-verb agreement	2L1 = MON	2L1 = eL2	eL2 < MON
	4	Verb meaning	2L1 = MON	2L1 = eL2	eL2 < MON
Late	5	Complex sentences	2L1 = MON	2L1 = eL2	eL2 < MON
	6	Wh-questions	2L1 = MON	2L1 = eL2	eL2 < MON
	6	Negation	2L1 < MON	2L1 = eL2	eL2 < MON
Very late	> 6	Case marking	2L1 < MON	2L1 = eL2	eL2 < MON

>: sig better than, <: sig worse than, =: n.s., shading indicates that this group has reached mastery in this scale at this test round.



the 2L1 and the monolingual group have not yet mastered Wh-questions, Negation and Case marking by age 4;4. Moreover, the 2L1 children perform better than the eL2 children in Subject-verb agreement, Verb meaning, Complex sentences and Wh-questions, but like the eL2 children on Negation and Case marking. By age 4;4 the eL2 group has not mastered any of the six phenomena.

As shown in **Table 9**, at age 5;8 the 2L1 children perform like their same-aged monolingual peers in Subject-verb agreement, Verb meaning, Complex sentences and Wh-questions, but worse than the monolinguals in Negation and Case marking. Regarding mastery, 2L1 children master Subject-verb agreement, Verb meaning and Complex sentences just like their monolingual peers. Unlike their monolingual peers, at age 5;8 2L1 children have not yet mastered Wh-questions and Negation. Both the 2L1 and the monolingual group have not yet mastered Case marking at age 5;8. Moreover, at this age the 2L1 children perform like the eL2 children in all six phenomena. Notably, at 5;8 years, the eL2 group has only mastered Verb meaning.

## DISCUSSION

In this paper we present data from children acquiring German as one of two languages or as the only language. Our goal was to investigate the effects of age of onset (from birth, around age 3) and of timing in L1 acquisition (early, late, very late) in child bilingualism. Three groups of children were included in the study: simultaneous bilingual children (2L1) and early second language learners of German (eL2) as well as monolingual children (MON). 2L1 children had an age of onset of 3 months, and the eL2 children had an age of onset of 35 months. To assess the stability of patterns across development, we collected data at two test rounds: at the age of 4;4 and about 16 months later. Crucially, the three groups did not differ in age (mean age 4;4 at test round 1 and 5;8 at test round 2). To study the factor timing in L1 acquisition, we targeted six phenomena that differ regarding the time at which they are mastered in monolingual acquisition: Subject-verb agreement (early), Verb meaning (early), Complex sentences (late), Wh-questions (late), Negation (late), and Case marking (very late). These phenomena were assessed using the standardized test LiSe-DaZ (Schulz and Tracy, 2011). Use of a standardized test has the advantage that timing in L1 acquisition could be determined by consulting the norming data for the monolingual children, independently from our sample but based on the same tasks as in the current study.

First we wanted to understand how the factors age of onset and timing in L1 acquisition affect the performance of simultaneous bilingual and early second language learning children. More specifically, we assessed the extent to which this difference in age of onset (from birth, around age 3) accounts for children's performance and whether timing in L1 acquisition interacts with this factor "age of onset." The differences and parallels found at age 4;4 between 2L1 and monolingual children on the one hand and between 2L1 and eL2 children on the other (see **Table 8**) point to a unique profile of simultaneous bilingual learners. This cannot be explained by age of onset alone, but

by the mediating effect of timing in L1 acquisition on age of onset. First, 2L1 children have more difficulty with later acquired phenomena than the same-aged monolingual children, despite their very similar age of onset and length of exposure at the time of testing. Second, for the (very) late acquired phenomena Negation and Case marking, the difficulties of the 2L1 learners are so prevalent that they perform on a par with eL2 learners. This is remarkable given that the simultaneous bilingual group had an age of onset of about 3 months, resulting in 50 months of exposure to German up to the age at testing, compared to an age of onset of 35 months in the early L2 group, resulting in only 16 months of exposure.

Put differently, at test round 1 the factor age of onset (from birth, around age 3) accounts for the consistent differences between eL2 and monolingual learners and for the observed partial differences between 2L1 and eL2 learners as well as for the observed partial parallels between 2L1 and monolingual learners. Timing in L1 acquisition accounts for the remaining patterns that would otherwise be unexpected: 2L1 learners show parallels to the eL2 learners and differences to the monolingual peers regarding the (very) late acquired phenomena Negation and Case marking. Furthermore, an intermediate pattern regarding Complex sentences and Wh-questions ( $MON > 2L1 > eL2$ ) suggests that the factor timing in L1 acquisition is sensitive enough to capture the difference between late phenomena to be mastered shortly after the age of testing in L1 acquisition and very late phenomena. Taken together, in (very) late acquired phenomena the factor age of onset is modulated by the factor timing in L1 acquisition.

The data of test round 2, collected 16 months after test round 1 (see **Table 9**), confirms the proposal suggested for the data of test round 1. At age 5;8 ceiling performance across all three groups was found only for the early acquired phenomenon Verb meaning. This means that for Verb meaning another 16 months of exposure to German were sufficient for the eL2 group to catch up to their simultaneous bilingual and monolingual peers, who had already been at ceiling at test round 1. Parallel to the pattern at test round 1 the 2L1 group performed worse than the monolingual group regarding the (very) late acquired phenomena Negation and Case marking at test round 2. This suggests that the simultaneous bilingual group did not profit from their early age of onset across all phenomena. What is more, in these (very) late acquired phenomena Negation and Case marking, the 2L1 group behaved just like the eL2 group. This finding supports the assumption that the simultaneous bilingual learners' earlier age of onset and considerably more exposure to German did not result in a general advantage over the early second language learners. For Wh-questions we found an intermediate pattern with the 2L1 group performing on a par with the monolinguals and with the eL2 group. These results clearly indicate that age of onset alone cannot account for the acquisition patterns found in our data. Timing in L1 acquisition contributes substantially to accounting for the observed parallels and differences between 2L1 and eL2 children on the one hand and 2L1 and monolingual children on the other. Note that the role of timing in L1 acquisition is most clearly visible in the (very) late acquired phenomena Complex sentences, Wh-questions,

Negation, and Case marking, which were tested *before* their mastery in monolingual children.

Our findings on age of onset effects in eL2 acquisition are in line with the research discussed earlier (see Section “Age of Onset Effects in Child Bilingualism”). Compared to monolingual children eL2 learners show a delay in rate of acquisition. Moreover, our data show that “catching up” does not necessarily happen at the same pace. Although both verb meaning and subject-verb agreement are acquired early, by age 5;8 eL2 children caught up to their monolingual peers in the former but not in the latter domain.

Crucially, our results indicate that simultaneous bilingual children do not consistently show the same acquisition rate and age of mastery as their monolingual peers. This contrasts with previous findings by Tracy and colleagues (Gawlitsek-Maiwald and Tracy, 1996; Tracy, 1995, unpublished) but confirms studies pointing to partial delays exhibited by 2L1 children, at least in one of their languages (Gathercole and Thomas, 2009; Bialystok and Hakuta, 2010). More specifically, our findings provide further evidence that age of onset effects are modulated by effects of timing of phenomena in L1 acquisition, as proposed by Tsimpli (2014). The design of our study enabled us to advance the debate on effects of timing in L1 acquisition in several ways. First, we assessed in the same groups of children a variety of phenomena in morpho-syntax and semantics differing with regard to timing in L1 acquisition. This permitted us to explore multiple asymmetries across domains. Second, using a standardized test (LiSe-DaZ, Schulz and Tracy, 2011) we could derive the ages of mastery directly from the monolingual norming sample. The resulting classification as early, late or very late acquired – based on the same tasks we employed with the participants in the current study – has the advantage of being well-defined and specific to the task. This freed us from the necessity of inferring information about age of mastery from the literature, which is likely to vary with the task used (see, for instance, Roesch and Chondrogianni, 2016, who argue based on production studies that case is early acquired in German). Third, since the children were assessed twice over an interval of 16 months, we could explore how parallels and differences between simultaneous bilingual children, early second language learners and monolingual children develop over time.

By comparing 2L1 and monolingual children as well as 2L1 and eL2 children, the present study extends previous research on late acquired phenomena such as gender (for Welsh: Gathercole and Thomas, 2009; for Dutch: Unsworth et al., 2014), comprehension of passives (Chondrogianni and Marinis, 2011; Armon-Lotem et al., 2015) and comprehension of *wh*-questions (Chondrogianni and Marinis, 2011; Roesch and Chondrogianni, 2016). In line with those studies, for the (very) late acquired phenomena sentential negation and case marking we did not find an effect of age of onset for the 2L1 group, i.e., we did not find an advantage of the 2L1 over the eL2 children. Rather, the 2L1 group, with an age of onset of about 3 months, performed as low as the eL2 group, who had an age of onset of about 35 months, on negation and case marking at both ages 4;4 and 5;8. Furthermore, in line with the timing hypothesis, the four late or very late acquired phenomena complex sentences, *wh*-questions, sentential negation, and case

marking were all found to pose difficulties for the simultaneous bilingual children. Despite their similar age of onset, the 2L1 children had significantly more difficulty than the monolingual children in these four phenomena at age 4;4. Notably, for negation and case marking this difference between 2L1 and monolingual children was still observable at age 5;8.

The second question we asked was whether language dominance influences bilingual children's performance across language domains. We addressed this question by looking more closely into the language dominance of the 2L1 group measured as language use at home; the majority of the eL2 children was dominant in the “other” language, as expected. Based on a parental questionnaire all 2L1 children were categorized as German-dominant ( $n = 18$ ), balanced ( $n = 12$ ) or non-German-dominant ( $n = 11$ ), depending on whether father, mother and siblings used German and/or the “other” language when speaking with the child. Except for negation, which was in fact easier for the non-German-dominant children than for the other two subgroups, there were no significant differences between the three groups at either age 4;4 or age 5;8.<sup>9</sup> This result was as expected for early acquired phenomena. Regarding late acquired phenomena the lack of advantage for German-dominant over non-German dominant simultaneous bilinguals was unexpected.

This missing effect of language dominance, measured as the dominant language used at home, for the simultaneous bilingual children in our study contrasts with studies that reported an effect of input factors on children's performance, such as amount of input (for vocabulary and morpho-syntax: Thordardottir, 2010; Paradis et al., 2011b; Hoff et al., 2012; for gender in Dutch: Unsworth, 2013b) and dominance (for vocabulary: Hammer et al., 2008). Our results agree with the findings by Chondrogianni and Marinis (2011) on eL2 children, who did not find amount of language input at home in the majority language to be related to children's language performance. The authors argue that their finding may be related to parents' low proficiency level in the majority language that the children were tested in (see also Paradis, 2011; Paradis and Jia, 2017). In our case, use of German as the dominant home language with the child did not facilitate simultaneous bilingual children's performance in either the early or the late acquired phenomena. This finding points to the general issue of how to assess the quality and native-likeness of the parental input that children are exposed to. Given that parental fluency has been found to be modulated by parental education (Chondrogianni and Marinis, 2011; Hoff et al., 2012), we may ask whether parental fluency and parental education in both children groups differed. While we could not collect data on the parents' level of proficiency or fluency to address this question directly, we do have information about parents' educational background. This data indicates that the parents of the German-dominant and non-German-dominant 2L1 children

<sup>9</sup>As pointed out by one of the reviewers, it comes as a surprise that German-dominant children were found to perform worse than non-German-dominant children at all. Note that this effect was present in one subtest, Negation, and at one test round only. We speculate that the way sentential negation was tested in the present study draws on comprehension abilities that are similar cross-linguistically, rendering dominance in the language of testing less important. Further studies are needed to explore this possibility.

had a similar educational background (see Footnote 8). It is hence likely that parental fluency in the tested majority language in both groups was similar as well. If parents' proficiency level in their L2 was low, then this factor may have overridden potential effects of language dominance, as the findings Chondrogianni and Marinis (2011) would suggest.

Alternatively, the presence of another developing language system may cause the language learner to weigh the two systems, which leads to the unique profile of simultaneous bilingual language learners. We agree with Tsimpli (2014) that timing of a structure in L1 acquisition is relevant, with late acquired phenomena being more difficult for 2L1 children than for their monolingual peers, making the 2L1 children look like early second language learners. However, rather than attributing these patterns to overall effects of input (e.g., in terms of length of exposure), we suggest switching perspective and looking at the acquisition task from the learner's point of view in order to capture the role of the input to the child in a more fine-grained way. We call this concept the learner's "internal need for time" to acquire a structure or property. More specifically, we propose that the amount of internal time needed is determined by two factors that have figured prominently in recent acquisition research: the "complexity" of the structure or property to be acquired and the "cross-linguistic robustness" of the phenomenon as well as the rule governing it. As noted before (Section "Timing in L1 Acquisition"), formal complexity may refer to how much idiosyncrasy and irregularity is involved in a construction (see Culicover, 2014), which may or may not align with the distinction between core and non-core. Cross-linguistic robustness refers to the issue of how much language-specific variation a construction or its interpretation exhibits (see e.g., the COST Action A 33 on cross-linguistically robust stages of children's linguistic performance). Put differently, we suggest that when complexity and cross-linguistic robustness are considered, the role of age of onset and language input for bilingual children's rate and success of acquisition can be addressed more comprehensively. A case in point are phenomena at the semantic-syntactic interface such as sentential negation tested in the current study. Sentential negation involves a complex mapping of syntactic position and meaning and is acquired late. Notably, the interpretation rules are assumed to be cross-linguistically the same, which arguably follows from the general assumption that well-formedness conditions on semantic representations are universal (see also Tsimpli, 2014: 296). Acquisition of sentential negation in bilingual children should hence be unsusceptible to differences in amount of input. This would also be compatible with the finding that in this scale German-dominant children actually performed worse than non-German-dominant children. The same reasoning holds for example for exhaustivity in single and multiple *wh*-questions, which is acquired late and seems to follow universal interpretation rules (Schulz, 2015). The situation is different for the German case marking paradigm tested in the current study. Because of its complex, intransparent form-function mapping described above, it is acquired very late. However, unlike sentential negation the system of case marking widely varies across languages, just like grammatical gender discussed above. Acquisition of these phenomena, which underlie

language-specific licensing rules, should be more sensitive to input effects. This is because, in addition to the time needed to weigh the two developing language systems, the more input the learner receives in the target-language the faster she can make the necessary language-specific choices. This proposal makes specific predictions based on cross-linguistic robustness that need to be tested in future studies.<sup>10</sup>

In this study we explored the interaction of age of onset (from birth, around age 3) and timing in L1 acquisition across different language domains and across development in two different groups of bilingual children that have been argued to constitute distinct acquisition types on theoretical grounds. Whereas chronological age and a number of external variables were controlled for, it was not possible to clearly dissociate age of onset effects from effects of length of exposure. In future research, a group of eL2 children who have the same length of exposure as the 2L1 group could be included as well as a bilingual sample in which age of onset is varied. Due to the set-up of the project, bilingual children acquired many different "other" languages; hence specific effects of the L1 could not be studied. Furthermore, as our small-scale longitudinal design revealed, time of testing plays an important role for detecting effects of both age of onset and timing in L1 acquisition. In future studies, the longitudinal aspect could be expanded, also assessing a wider range of late acquired phenomena. Use of a standardized test allowed us to assess a number of different phenomena and to derive precise ages of mastery. Future experimental studies targeting the phenomena in more detail could shed light on the variation within a scale, e.g., dative case being acquired later than accusative case. Our finding that language dominance did not affect 2L1 children's performance could be followed up with studies employing more fine-grained measures of dominance including assessment of parental language proficiency. Finally, 2L1 children were found to perform better on the non-verbal IQ test than both the monolingual and the eL2 children. However, the role of non-verbal IQ was limited to few subscales and did not reveal any systematic pattern. This result is in line with previous studies (Schulz and Tracy, 2011: 109; Wojtecka, 2018, unpublished) that found only very few, weak correlations between non-verbal IQ and performance on the LiSe-DaZ subscales. Future studies could explore this factor in more detail for the group of 2L1 children.

## CONCLUSION

Our findings indicate that in the context of children acquiring German, timing in L1 acquisition is an important factor in child bilingual acquisition, interacting with effects of age of onset, even for learners with an initial exposure to German before age four. Whereas bilingual children's performance in early acquired phenomena could be explained by age of onset effects alone,

<sup>10</sup>Thanks to C. Hamann for pointing out to us that lateness of negation could also be language-specific. The interplay of syntactic position and meaning may be particularly difficult for German negation because of its less obvious scope properties. Further studies are needed to clarify this point.

only the impact of timing could account for pace and success of acquisition in late acquired phenomena. The observation that the factor language dominance for the simultaneous bilingual group, measured as language use at home, did not affect children's rate of acquisition, led us to propose an alternative concept to capture the apparent role of input: the learner's need for time to master a phenomenon, which is determined by its complexity and by its cross-linguistic robustness.

## ETHICS STATEMENT

The study was based on data from two projects (MILA and cammino), both conducted in accordance with the Declaration of Helsinki; informed written consent was obtained from the parents of all participants, including publication of the results. The project MILA (PI: PS) was approved by the ethics committee of the German Psychological Association (DGPs) on March 12, 2009. The project cammino (PIs: PS, AG) was approved by the ethics committee of the Department of Psychology and Sports Sciences of Goethe University on June 24, 2011.

## AUTHOR CONTRIBUTIONS

PS contributed conception and design of the study and wrote Sections "Introduction," "Factors Influencing Child Bilingual Acquisition," "Results," and "Discussion." AG was responsible for managing data compilation and performing the statistical analysis, and wrote Section "Materials and Methods." Both

authors contributed to manuscript revision of the text and read and approved the submitted version.

## FUNDING

The research presented here was part of the projects MILA (PI: PS) and cammino (PIs: PS, AG) and was carried out at the Research Center IDEA. MILA was funded from 2008 to 2014 by the LOEWE program for Excellency from the State of Hesse. Cammino (01NV1011 und 01NV1012) was funded from 2011 to 2015 by the Federal Ministry for Research and Education (BMBF) in the research area "Cooperation in elementary and primary education."

## ACKNOWLEDGMENTS

We thank Barbara Geist, Rabea Lemmer, Barbara Voet Cornelli, and Magda Wojtecka for their help with managing the research projects and Daniel Liebner for statistical support. We are grateful to our research assistants, to the children who participated and their parents, and to the teachers in the kindergartens for their support. Previous versions of this study have been presented at Bi-SLI Tours 2015 and Bi-SLI Reading 2018. We are grateful to Ana Pérez-Leroux, Ianthi Tsimpli, Laurie Tuller, and Merle Weicker for helpful discussion and to Jan-Henning Ehm for statistical advice. The comments of the reviewers and the editors have greatly helped to improve the quality of the manuscript.

## REFERENCES

- Armon-Lotem, S., Haman, E., Jensen de López, K., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., et al. (2015). A large-scale cross-linguistic investigation of the acquisition of passive. *Lang. Acquis.* 23, 27–56. doi: 10.1080/10489223.2015.1047095
- Bialystok, E. (2009). Bilingualism: the good, the bad, and the indifferent. *Bilingualism* 12, 3–11. doi: 10.1017/S1366728908003477
- Bialystok, E., and Hakuta, K. (2010). "Confounded age: linguistic and cognitive factors in age differences for second language acquisition," in *Second Language Acquisition and the Critical Period Hypothesis*, ed. D. Birdsong. (New York, NY: Routledge), 161–182.
- Birdsong, D. (2006). Age and second language acquisition and processing: a selective overview. *Lang. Learn.* 56, 9–49. doi: 10.1111/j.1467-9922.2006.00353.x
- Blom, E., Polišká, D., and Unsworth, S. (2008). The acquisition of grammatical gender in Dutch. *Second Lang. Res.* 24, 259–265. doi: 10.1177/0267658308090187
- Bloom, L., Rispoli, M., Gartner, B., and Hafitz, J. (1989). Acquisition of complementation. *J. Child Lang.* 16, 101–120. doi: 10.1017/S0305000900013465
- Bohman, T. M., Bedore, L. M., Peña, E. D., Mendez-Perez, A., and Gillam, R. B. (2010). What you hear and what you say: language performance in Spanish-English bilinguals. *Int. J. Biling. Educ. Biling.* 13, 325–344. doi: 10.1080/13670050903342019
- Chilla, S. (2008). *Erstsprache, Zweitsprache, spezifische Sprachentwicklungsstörung?: Eine Untersuchung des Erwerbs der deutschen Hauptsatzstruktur durch sukzessiv-bilinguale Kinder mit türkischer Erstsprache*. Hamburg: Kovač.
- Chondrogianni, V. (2018). "Child L2 acquisition," in *Bilingual Cognition and Language: The State of the Science Across Subfields*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins), 103–126. doi: 10.1075/sibil.54.06cho
- Chondrogianni, V., and Marinis, T. (2011). Internal and external factors in child second language acquisition. *Ling. Approaches Biling.* 1, 318–345. doi: 10.1075/lab.1.3.05cho
- Clahsen, H. (1986). *Die Profilanalyse: Ein linguistisches Verfahren für die Sprachdiagnose im Vorschulalter*. Berlin: Marhold.
- Culicover, P. W. (2014). "Constructions, complexity, and word order variations," in *Measuring Grammatical Complexity*, eds F. J. Newmeyer and L. B. Preston. Oxford: Oxford University Press, 148–178.
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Bristol: Multilingual Matters.
- Eichler, N., Jansen, V., and Müller, N. (2013). Gender acquisition in bilingual children: French–German, Italian–German, Spanish–German and Italian–French. *Int. J. Biling.* 17, 550–572. doi: 10.1177/1367006911435719
- Eisenbeiss, S., Bartke, S., and Clahsen, H. (2005). Structural and lexical case in child German: evidence from language-impaired and typically-developing children. *Lang. Acquis.* 13, 3–32. doi: 10.1207/s15327817la1301\_2
- Field, A. (2013). *Discovering Statistics Using SPSS: (and Sex and Drugs and Rock 'n' Roll)*. Los Angeles, CA: Sage.
- 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
- Gathercole, V. C. M., and Thomas, E. M. (2009). Bilingual first-language development: dominant language takeover, threatened minority language take-up. *Bilingualism* 12, 213–237. doi: 10.1017/S1366728909004015
- Gawltzek-Maiwald, I., and Tracy, R. (1996). Bilingual bootstrapping. *Linguistics* 34, 901–926. doi: 10.1515/ling.1996.34.5.901



- Genesee, F., and Nicoladis, E. (2007). "Bilingual first language acquisition," in *Blackwell handbook of language development*, eds E. Hoff & M. Shatz (Malden, MA: Blackwell), 324–342. doi: 10.1002/9780470757833.ch16
- Grimm, A., and Schulz, P. (2012). "Forschungsmethoden der kombinierten Längs- und Querschnittstudie MILA," in *DaZ-Forschung: Deutsch als Zweitsprache, Mehrsprachigkeit und Migration*, eds B. Ahrenholz, C. Dimroth, B. Lütke, & M. Rost-Roth (Berlin: De Gruyter), 195–218.
- Grimm, A., and Schulz, P. (2014a). 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
- Grimm, A., and Schulz, P. (2014b). "Sprachfähigkeiten von Kindern mit DaZ bei Schuleintritt," in *Deutsch als Zweitsprache: erwerben, lernen und lehren: Beiträge zum 9. Workshop Kinder mit Migrationshintergrund*, eds B. Lütke & I. Petersen (Stuttgart: Klett Fillibach), 35–50
- Grimm, A., and Schulz, P. (2016). Warum man bei mehrsprachigen Kindern dreimal nach dem Alter fragen sollte: Sprachfähigkeiten simultan-bilingualer Lerner im Vergleich mit monolingualen und frühen Zweitsprachlernern. *Diskurs Kindheits- Und Jugendforschung*, 11, 27–42. doi: 10.3224/diskurs.v11i1.22247
- Haberzettl, S. (2005). *Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache*. Tübingen: Niemeyer. doi: 10.1515/9783110922127
- Hager, M., and Müller, N. (2015). Ultimate attainment in bilingual first language acquisition. *Lingua* 164, 289–308. doi: 10.1016/j.lingua.2015.05.014
- Hammer, C. S., Lawrence, F. R., and Miccio, A. W. (2008). Exposure to english before and after entry into head start: bilingual children's receptive language growth in Spanish and English. *Int. J. Biling. Educ. Biling* 11, 30–56. doi: 10.2167/beb376.0
- 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
- Hopp, H. (2011). Internal and external factors in child second language acquisition. *Linguist. Approaches Biling* 1, 238–264. doi: 10.1075/lab.1.3.02hop
- Johnson, J. S., and Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cogn. Psychol.* 21, 60–99. doi: 10.1016/0010-0285(89)90003-0
- Kaufman, A. S., Kaufman, N. L., and Melchers, P. (2003). *Kaufman Assessment Battery for Children: K-ABC*. Leiden: PITS.
- Klein, W. (1996). Language acquisition at different ages. In D. Magnusson (Ed.), *The Lifespan Development of Individuals: Behavioral, Neurobiological, and Psychosocial Perspectives. A Synthesis*. Cambridge: Cambridge University Press, 244–264
- Kupisch, T. (2006). *The Acquisition of Determiners in Bilingual German-Italian and German-French Children*. München: Lingcom Europa.
- Kupisch, T., and Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Biling.* 22, 564–582. doi: 10.1177/1367006916654355
- Locke, J. L. (1997). A theory of neurolinguistic development. *Brain Lang.* 58, 265–326. doi: 10.1006/brln.1997.1791
- Meisel, J. M. (2008). Child second language acquisition or successive first language acquisition? in *Current Trends in Child Second Language Acquisition: A Generative Perspective*, eds B. Haznedar & E. Gavruseva (Amsterdam: John Benjamins), 55–82. doi: 10.1075/lald.46.04mei
- Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift Für Sprachwissenschaft* 28, 5–34. doi: 10.1515/ZFSW.2009.002
- Meisel, J. M. (2011). *First and Second Language Acquisition: Parallels and Differences*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511862694
- Meisel, J. M. (2013). "Sensitive phases in successive language acquisition: the critical period hypothesis revisited," in *The Cambridge handbook of biolinguistics*, ed. C. Boeckx, K. K. Grohmann. Cambridge: Cambridge University Press. 69–85
- Meisel, J. M. (2016). Early child second language acquisition: french gender in German children. *Bilingualism* 28, 1–18. doi: 10.1017/S1366728916000237
- Müller, N. (1990). Erwerb der Wortstellung im Französischen und Deutschen. Zur Distribution von Finitivmerkmalen in der Grammatik bilingualer Kinder. *Linguistische Berichte*. 3, 127–151. doi: 10.1007/978-3-663-14309-3\_6
- Müller, N. (2017). Different sources of delay and acceleration in early child bilingualism. *Zeitschrift Für Sprachwissenschaft* 36, 7–30. doi: 10.1515/zfs-2017-0002
- Newmeyer, F. J., and Preston, L. B. (Eds.). (2014). *Measuring Grammatical Complexity*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199685301.001.0001
- Paradis, J. (2011). Internal and external factors in child second language acquisition. *Linguist. Approaches Biling.* 1, 213–237. doi: 10.1075/lab.1.3.01par
- Paradis, J., and Genesee, F. (1996). Syntactic acquisition in bilingual children. *Stud. Second Lang. Acquisit.* 18, 1–25. doi: 10.1017/S0272263100014662
- Paradis, J., Genesee, F., & Crago, M. B. (Eds.). (2011a). *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*, 2nd Edn. Baltimore, MD: Brookes.
- 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., Nicoladis, E., Crago, M. B., and Genesee, F. (2011b). Bilingual children's acquisition of the past tense: a usage-based approach. *J. Child Lang.* 38, 554–578. doi: 10.1017/S0305000910000218
- Penner, Z., Schulz, P., and Wymann, K. (2003). Learning the meaning of verbs: what distinguishes language impaired from normally developing children? *Linguistics* 41, 289–319. doi: 10.1515/ling.2003.010
- Prévost, P. (2003). Truncation and missing surface inflection in initial child L2 German. *Stud. Second Lang. Acquisit* 25, 65–97. doi: 10.1017/S0272263103000032
- Roeper, T., Schulz, P., Pearson, B. Z., and Reckling, I. (2007). "From singleton to exhaustive: the acquisition of Wh-," in *Proceedings of the 3rd Conference on the Semantics of Underrepresented Languages in the Americas*, eds M. Becker & A. McKenzie (Amherst, MA: Book Surge Publishing), 87–102.
- Roesch, A. D., and Chondrogianni, V. (2016). "Which mouse kissed the frog?": effects of age of onset, length of exposure, and knowledge of case marking on the comprehension of wh-questions in German-speaking simultaneous and early sequential bilingual children. *J. Child Lang.* 43, 635–661. doi: 10.1017/S0305000916000015
- Rothman, J., Long, D., Iverson, M., Judy, T., Lingwall, A., and Chakravarty, T. (2016). Older age of onset in child L2 acquisition can be facilitative: evidence from the acquisition of English passives by Spanish natives. *J. Child Lang.* 43, 662–686. doi: 10.1017/S0305000915000549
- Rothweiler, M. (2006). "The acquisition of V2 and subordinate clauses in early successive acquisition of German," in *Interfaces in Multilingualism: Acquisition and Representation*, ed. C. Lleó (Amsterdam: John Benjamins), 91–113.
- Rothweiler, M. (2007). "Bilingualer Spracherwerb und Zweitspracherwerb," in *Schnittstellen der germanistischen Linguistik*, ed. M. Steinbach (Stuttgart: Metzler), 103–130.
- Rothweiler, M., Schönenberger, M., and Sterner, F. (2017). Subject-verb agreement in German in bilingual children with and without SLI. *Zeitschrift Für Sprachwissenschaft* 36, 79–106. doi: 10.1515/zfs-2017-0005
- Schmeißer, A., Hager, M., Gil, L. A., Jansen, V., Geveler, J., Eichler, N., Müller, N. (2015). "Related but different: the two concepts of language dominance and language proficiency," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 36–65. doi: 10.1017/CBO9781107375345.003
- Schulz, P. (2007). "Erstspracherwerb Deutsch: Sprachliche Fähigkeiten von eins bis zehn," in *Diagnostik und Förderung im Elementarbereich und Grundschulunterricht: Lernprozesse wahrnehmen, deuten und begleiten*, eds U. Graf and E. Moser Opitz (Baltmannsweiler: Schneider Verlag Hohengehren), 67–85
- Schulz, P. (2012). Why and how individual differences matter for linguistic theory and experimental research but not for UG. *Linguist. Approaches Biling.* 2, 298–303. doi: 10.1075/lab.2.3.11sch
- Schulz, P. (2013). "Wer versteht wann was? Sprachverstehen im frühen Zweitspracherwerb des Deutschen am Beispiel der w-Fragen," in *Das Deutsch der Migranten*, ed. A. Deppermann (Berlin: De Gruyter), 313–337. doi: 10.1515/9783110307894.313
- Schulz, P. (2015). "Comprehension of wh-questions," in *Assessing Multilingual Children: Disentangling Bilingualism From Language Impairment*, eds S.

- Armon-Lotem, J. D. Jong, and N. Meir (Bristol: Multilingual Matters), 76–94. doi: 10.21832/9781783093137-006
- Schulz, P. (2018). Acquisition of telicity. in *Semantics in Language Acquisition* eds K. Syrett & S. Arunachalam (Amsterdam: John Benjamins).
- Schulz, P., and Grimm, A. (2012). “Spracherwerb,” in *Germanistik: Sprachwissenschaft – Literaturwissenschaft – Schlüsselkompetenzen*, eds H. Drügh and H.-H. Ewers (Stuttgart: Metzler), 155–172. doi: 10.1007/978-3-476-00399-7\_6
- Schulz, P., Grimm, A., Geist, B., and Voet Cornelli, B. (2014). “Cammino - Mehrsprachigkeit Am Übergang zwischen Kita und Grundschule,” in *Bildungsforschung 2020 - Herausforderungen und Perspektiven. Dokumentation der Tagung des Bundesministeriums für Bildung und Forschung vom 29.-30. März 2012*, eds Bundesministerium für Bildung und Forschung (Bonn: Bundesministerium für Bildung und Forschung (BMBF)), 281–284.
- 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 Schwarze, R. (2017). How strong is the ban on non-finite verbs in V2? Evidence from early second language learners of German with and without SLI. *Zeitschrift Für Sprachwissenschaft* 36, 51–78. doi: 10.1515/zfs-2017-0004
- Schulz, P., and Tracy, R. (2011). *LiSe-DaZ - Linguistische Sprachstandserhebung Deutsch als Zweitsprache*. Göttingen: Hogrefe.
- Schwartz, B. (2004). On child L2 development of syntax and morphology. *Lingue E Linguaggio* 3, 97–132.
- Schwarze, R. (2018). *Sprachentwicklungsstörungen bei frühen Zweitsprachlernern: Der Erwerb von Kasus, Finitheit und Verbstellung*. Available at: <http://publikationen.uni-frankfurt.de/frontdoor/index/index/docId/46583>
- Thordardottir, E. (2010). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 15, 426–445. doi: 10.1177/1367006911403202
- Tracy, R. (1986). The acquisition of case morphology in German. *Linguistics* 24, 47–78. doi: 10.1515/ling.1986.24.1.47
- Tracy, R. (1991). *Sprachliche Strukturentwicklung: Linguistische und kognitionspsychologische Aspekte einer Theorie des Erstspracherwerbs*. Tübingen: Narr.
- Tracy, R., and Lemke, V. (2012). “Young L2 and Young L2 and L1 Learners: more alike than different,” in *Second Language Acquisition: Comparative Perspectives on Language Acquisition: A Tribute to Clive Perdue*, Vol. 61, ed M. Watorek (Bristol: Multilingual Matters), 303–323.
- Tracy, R., and Thoma, D. (2009). “Convergence on finite V2 clauses in L1, bilingual L1 and early L2 acquisition,” in *Functional Categories in Learner Language*, eds C. Dimroth & P. Jordens (Berlin: De Gruyter), 1–44.
- Tsimpli, I. M. (2003). Features in L1 and L2 acquisition: evidence from Greek clitics and determiners. *Acquisition Et Interaction En Langue Etrangere* 20, 87–128.
- Tsimpli, I. M. (2014). Early, late or very late? *Linguist. Approaches Biling.* 4, 283–313. doi: 10.1075/lab.4.3.01tsi
- 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. D. Jong, & N. Meir (Bristol: Multilingual Matters), 301–330. doi: 10.21832/9781783093137-013
- Unsworth, S. (2013a). Assessing age of onset effects in (early) child L2 acquisition. *Lang. Acquisit.* 20, 74–92. doi: 10.1080/10489223.2013.766739
- Unsworth, S. (2013b). 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. (2016). Early child L2 acquisition: age or input effects? Neither, or both? *J. Child Lang.* 43, 608–634. doi: 10.1017/S030500091500080X
- Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., and Tsimpli, I. M. (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
- Wegener, H. (1998). “Das Passiv im DaZ-Erwerb von Grundschulkindern,” in *Eine zweite Sprache lernen: Empirische Untersuchungen zum Zweitspracherwerb*, ed. H. Wegener (Tübingen: Narr), 143–172.
- Wojtecka, M., Koch, C., Grimm, A., and Schulz, P. (2011). “Production and comprehension of sentence negation in child German,” in *Production-Comprehension Asymmetries in Child Language*, eds A. Grimm, A. Müller, C. Hamann, & E. Ruigendijk (Berlin: De Gruyter), 217–246.

**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 © 2019 Schulz and Grimm. 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.



# Experiential Measures Can Be Used as a Proxy for Language Dominance in Bilingual Language Acquisition Research

Sharon Unsworth<sup>1\*</sup>, Vicky Chondrogianni<sup>2</sup> and Barbora Skarabela<sup>2</sup>

<sup>1</sup> Centre for Language Studies, Radboud University, Nijmegen, Netherlands, <sup>2</sup> School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh, United Kingdom

## OPEN ACCESS

### Edited by:

Cornelia Hamann,  
University of Oldenburg, Germany

### Reviewed by:

István Fekete,  
University of Oldenburg, Germany  
Theres Gruter,  
University of Hawaii, United States

### \*Correspondence:

Sharon Unsworth  
s.unsworth@let.ru.nl

### Specialty section:

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

**Received:** 26 April 2018

**Accepted:** 05 September 2018

**Published:** 17 October 2018

### Citation:

Unsworth S, Chondrogianni V and  
Skarabela B (2018) Experiential  
Measures Can Be Used as a Proxy  
for Language Dominance in Bilingual  
Language Acquisition Research.  
Front. Psychol. 9:1809.  
doi: 10.3389/fpsyg.2018.01809

Language dominance is a multidimensional construct comprising several distinct yet interrelated components, including language proficiency, exposure and use. The exact relation between these components remains unclear. Several studies have observed a (non-linear) relationship between bilingual children's amount of exposure and absolute proficiency in each language, but our understanding of the relationship between language exposure and use and relative proficiency is limited. To address this question, we examined whether experiential-based measures of language dominance, operationalised here in the narrow sense of relative language proficiency, can provide an efficient alternative to the more labor-intensive performance-based measures often used in the literature. In earlier work, Unsworth (2016a) examined the relationship between relative proficiency and language exposure and use in a group of English–Dutch bilingual preschool children residing in the Netherlands. This study expands these findings by examining Dutch–English preschool children of the same age residing in the United Kingdom in order to cover the full dominance continuum. Participants were 35 simultaneous bilingual children (2;0–5;0) exposed to English and Dutch, 20 resident in the Netherlands and 15 in the United Kingdom. Relative amount of language exposure and use were estimated using a parental questionnaire. To obtain performance-based measures of language proficiency, children's spontaneous speech was recorded during a half-hour play session in each language. The transcribed data were used to derive MLU (words), average length of the longest five utterances, the number of different verb and noun types. Single word vocabulary comprehension was assessed using standardized tests in both languages. Following Yip and Matthews (2006), relative proficiency was operationalised using differentials. In line with Unsworth (2016a), English-dominant children typically had less than approx. 35% exposure to Dutch and used Dutch less than approximately 30% of the time. Curve-fitting analyses revealed that non-linear models best fit the data. Logistic regression analyses showed that both exposure and use were good predictors of dominance group membership assigned using the same approach as Unsworth (2016a), that is, using SDs. Dominance groups derived independently using cluster analyses overlapped with the groups derived using SDs, confirming that relative amount of exposure and use can be used as a proxy for language dominance.

**Keywords:** language dominance, language exposure, language use, bilingual children, relative proficiency

## INTRODUCTION

Bilingual children's language development is affected by certain characteristics of their language learning experience. For example, numerous studies have found that bilingual children's rate of acquisition in vocabulary and grammar is often predicted by the relative amount of input to which they are exposed in their two languages (e.g., Hoff et al., 2012; see Unsworth, 2016b for a recent overview). Similarly, a number of recent studies have demonstrated that children's own language use has a significant impact on their language development across a range of domains (e.g., Bohman et al., 2010; Ribot et al., 2017). Almost all of these studies focus on the relationship between (relative) measures of language experience and children's *absolute* proficiency in one or both of their two languages. Considerably fewer studies (e.g., Bedore et al., 2012) have explored the relationship between these experiential variables and children's *relative* proficiency, that is, how well children perform in one language compared with the other.

Bilinguals who are equally proficient – or balanced – in their two languages are rare (e.g., Grosjean, 1982, 2010). Whilst there is a broad consensus that bilingual children, even if exposed to both languages from birth, are more proficient or *dominant* in one of their two languages, how best to define language dominance remains a contentious issue. Part of the challenge in defining language dominance is that it involves a multidimensional construct consisting of several distinct yet interrelated components, including language use, language input, and language proficiency (see Montrul, 2016 and Silva-Corvalán and Treffers-Daller, 2016 for relevant discussion). In the present study, we adopt a narrow definition of *language dominance*, focusing on relative proficiency across the two languages, in order to explore the relationship between these various component parts.

In the bilingual acquisition literature, dominance is typically operationalised using either *performance-based measures* such as mean length of utterance (MLU) or lexical diversity (e.g., Cantone et al., 2008) or *experiential-based measures* such as amount of exposure or country of residence (e.g., Döpke, 1992; Argyri and Sorace, 2007; Foroodi-Nejad and Paradis, 2009; Serratrice et al., 2009; Tuller et al., 2018). Whilst the former are arguably more objective, they are considerably more time-consuming and consequently more expensive than the latter; experiential-based measures are certainly cheaper and quicker to administer but they are often considered subjective and rather crude in comparison with performance-based measures. It is, however, unclear whether this is indeed the case.

The goal of the present study is, therefore, to assess whether and to what extent experiential-based measures can be used as a proxy for language dominance in bilingual language development research. The paper is organized as follows. First, we briefly summarize the main findings concerning the relationship between language exposure and use and *absolute* language proficiency. Next, we review the more limited previous literature examining the relationship between these two experiential variables and *relative* language proficiency, including a study by Unsworth (2016a) on English–Dutch bilingual preschoolers in the Netherlands, which serves as the starting point for the

present study. Subsequently, we combine data from this earlier study with new data from English–Dutch bilingual preschoolers in the United Kingdom in order to explore the relationship between language exposure and use and language dominance across the whole dominance continuum. Our main finding is that in line with Bedore et al. (2012) and Unsworth (2016a), there is a moderate to strong non-linear relationship between language exposure and use, on the one hand, and relative language proficiency, on the other, suggesting that experiential-based measures can indeed be used as a proxy for performance-based measures of language dominance.

## ABSOLUTE AND RELATIVE PROFICIENCY

### Language Exposure and Use and Absolute Proficiency

Bilingual children's language experience varies considerably. Whilst some children hear the minority language from both parents, others receive minority language input from one parent only. For some children, their parent(s) are the only source of the minority language, whereas others have access to minority language input from other family members and friends. Furthermore, some children hear language input from native speakers only, yet others also hear language input – sometimes exclusively so – from non-native speakers. There is also a difference in the availability of TV, apps and other media across different languages. Taken together, this variability in how bilingual children experience their two languages means that there is considerable variation between and sometimes also within children in terms of the quantity and quality of language exposure and use. This variation has been observed to predict bilingual children's developing language skills across a range of linguistic domains, language combinations and sociolinguistic settings.

Input quantity effects have been observed for a range of different domains of bilingual children's language proficiency, such as vocabulary (Gathercole and Thomas, 2009; Thordardottir, 2011), aspects of morphosyntax such as MLU (Place and Hoff, 2011; Hoff et al., 2012) and verbal morphology (Nicoladis et al., 2007; Blom, 2010; Paradis et al., 2011), as well as certain phonological abilities (Sundara et al., 2008; Nicoladis and Paradis, 2011). The relationship between relative amount of exposure and language skills has been found to be non-linear in nature. This means, for example, that once bilingual children reach a certain input threshold, they score on a par with monolingual peers but beyond that threshold, the relationship between exposure and proficiency is more limited, if present at all (e.g., Pearson et al., 1997; Cattani et al., 2014; Thordardottir, 2015). Differential effects of input have been observed for toddlers (e.g., Place and Hoff, 2011), preschoolers (e.g., Paradis et al., 2011) and primary school children (e.g., Gathercole and Thomas, 2009), and in both simultaneous (e.g., Unsworth, 2013) and successive (e.g., Chondrogianni and Marinis, 2011; Paradis, 2011) bilingual children, and in minority



language (e.g., Hoff et al., 2012) and bilingual (e.g., Gathercole and Thomas, 2009) sociolinguistic contexts. In short, there is considerable evidence for a robust relationship between amount of language exposure and rate of acquisition in bilingual language development. Most of the aforementioned studies concerned bilingual children who were still relatively young and therefore unlikely to have reached their end state in one or both of their two languages. In other words, most of the literature on input effects deals with *rate* of acquisition rather than the end state. It remains unclear whether amount of language exposure in early childhood is also a strong predictor of children's long-term outcomes (but see e.g., Montrul, 2008 for evidence that it likely is).

In addition to language exposure, children's own language use has also been found to play a significant role in their bilingual development. Several recent studies have shown that the extent to which children actively speak the language in question significantly predicts their developing language abilities. For example, a study on the early semantic and morphosyntactic development of Spanish-English bilinguals found that children's language use was a significant predictor of both domains in both languages, whereas input was only relevant for both domains in English (Bohman et al., 2010). Similar findings were reported for children's morphosyntactic development and vocabulary size (e.g., Montrul, 2008; Paradis, 2011; Hammer et al., 2012). More recently, Ribot et al. (2017) also observed that after controlling for input effects, language use at 30 months predicted bilingual Spanish-English children's expressive vocabulary skills at 36 and 42 months, but this was not the case for receptive skills. More specifically, children whose use in English was greater than their input in English (i.e., children who sometimes switched to English when spoken to in Spanish) had higher expressive vocabulary scores and their scores increased at a faster rate than children whose output in English was less than their input in English (i.e., children who sometimes switched to Spanish when spoken to in English). The effect of language use on children's Spanish skills was not assessed. To summarize, there is emerging evidence suggesting that in addition to language exposure, language use has also been found to predict unique variance in bilingual children's absolute proficiency in one or both of their languages.

## Language Exposure and Use and Relative Proficiency

As noted above, there are comparatively fewer studies examining the relationship between language exposure and use and *relative* rather than *absolute* language proficiency, even though relative language exposure is often used as a proxy for language dominance (e.g., Foroodi-Nejad and Paradis, 2009). It has been argued that using relative measures of exposure to predict absolute measures of language skill fails to capture variation in the overall amount of child-directed speech to which children are exposed (i.e., their absolute exposure), and that this variability can be controlled for by comparing relative measures of exposure with relative measures of proficiency, particularly when the goal is to better understand patterns of language balance or dominance (Grüter et al., 2014), as is the case here.

There are a number of ways in which absolute language proficiency scores in two languages can be combined to provide some measure of *relative* language proficiency or language dominance (in the narrow sense in which it is used here). One common approach in the bilingual language acquisition literature, which is also adopted here, is to use differentials (e.g., Yip and Matthews, 2006). Differentials involve subtracting a child's score on one language from his or her score on the other language; this method can in principle be adopted with any measure of language proficiency, although scores need to be standardized before subtraction if they are to be directly comparable across different measures (for relevant discussion see Birdsong, 2016 and Treffers-Daller and Korybski, 2016).

One of the few studies using exposure and use to predict relative proficiency is Bedore et al. (2012). A large sample of 5-year-old Spanish-English bilinguals ( $n = 1029$ ) participated in the semantics and morphosyntax subtests of the BESOS (Bilingual English Spanish Oral Language Screening, developed by the same authors), both productive tasks. Following the standard practice in the field, measures of language experience were estimated from parental questionnaires. Children were divided into dominance groups (i.e., functionally monolingual in Spanish, bilingual Spanish dominant, balanced bilingual, bilingual English dominant, functionally monolingual in English) based on experiential-based measures (i.e., differences in their relative language exposure and use in English and Spanish) and on performance-based measures (i.e., differences in their scores on the morphosyntactic and semantics subtests in English and Spanish). The authors observed that the dominance profiles derived from the two experiential-based measures were more consistent with each other than those derived from the two performance-based measures, although this is perhaps unsurprising given that language exposure and use were so highly correlated in their sample ( $r = 0.95$ ). A combined current usage score based on these two factors was found to account for more variance in children's relative morphosyntactic and semantic proficiency than age of first exposure. The authors concluded that current usage should therefore be included in assessing dominance patterns in (5-year-old) bilingual children. As the authors note, however, given that language use has been found to relate to language development differently from language input (Bohman et al., 2010), it is nevertheless important to consider the two separately.

In a smaller scale study with younger children, Unsworth (2016a) investigated the relationship between relative proficiency and language exposure, on the one hand, and language use, on the other. More specifically, Unsworth compared a number of performance-based commonly used measures of language dominance derived from spontaneous speech samples (i.e., MLU and various measures of lexical diversity) with experiential-based measures derived from parental questionnaires to explore whether the latter could reasonably be used as a proxy for the former. Participants were 18 simultaneous bilingual English-Dutch children aged between 2 and 4 years old. Despite being strongly correlated with each other ( $r = 0.80$ ), the children's patterns of language exposure and use related differently to their relative proficiency, as determined by differentials (following Yip

and Matthews, 2006). More specifically, whereas the children classified as Dutch-dominant on the basis of their differential scores all had at least 65% relative exposure to Dutch, they used Dutch almost exclusively, at least 90% of the time. On the basis of these findings, Unsworth concluded that, in line with previous work on the relationship between exposure and absolute proficiency (see above), experiential-based measures may be used as a proxy for language dominance.

The children in Unsworth's study were all resident in the Netherlands and were all found to be Dutch-dominant or balanced. To fully understand the potential of using amount of exposure as proxy for language dominance, it is important to include children from the whole dominance continuum. For this reason, the present study combines the original data in Unsworth (2016a) with new data from English–Dutch bilingual children resident in the United Kingdom to create a larger sample, including bilingual children in a primarily English-speaking environment, and allowing us to conduct a range of analyses. Our research questions are as follows:

- (1) To what extent do experiential-based factors such as relative language exposure and use predict bilingual children's relative language proficiency?
- (2) Can language exposure and use reliably classify bilingual children into language dominance groups?
- (3) Could experiential-based measures be used as a proxy for language dominance?

Based on previous work focusing on absolute proficiency, as well as Bedore et al. (2012) and Unsworth (2016a), we predict that relative proficiency scores will correlate strongly with both measures of language exposure and use. The exact nature of this relationship, however, will not necessarily be the same (e.g., Ribot et al., 2017) and it will not be linear (e.g., Thordardottir, 2011; Cattani et al., 2014). More specifically, assuming the United Kingdom-based children will pattern similarly to their peers in the Netherlands, we expect in answer to the second research question that children classified as English-dominant should have no more than 35% exposure to Dutch and no more than 10% of their language output in Dutch. Similarly, children classified as balanced bilinguals on the basis of their differential scores should hear and use Dutch more than the English-dominant children, but less than the values observed for Dutch-dominant children in the original study (i.e., below 65% for language exposure and 90% for use). If these expectations are borne out, then experiential-based measures of language exposure and use could arguably be used as a proxy for language dominance.

## MATERIALS AND METHODS

### Participants

Participants were 35 simultaneous bilingual children exposed to English and Dutch, 20 resident in the Netherlands (age range: 2;9 – 4;6;  $M = 3;9$ ;  $SD = 0;7$ ; 7 girls, 18 taken from Unsworth, 2016a plus two additional children) and 15 resident in the United Kingdom (age range: 2;0 – 5;1;  $M = 3;5$ ;  $SD = 1;1$ ; 10 girls). All but two children in the Netherlands and all but one child in the

United Kingdom were exposed to both languages from birth; the three exceptions were all exposed to both languages before the age of two. Their inclusion in the analyses did not affect any of the results.

The children in the Netherlands were almost all being raised following the *one parent, one language* approach: in twelve families the mother mostly or always spoke English to the child and the father mostly or always spoke Dutch, and in seven families this pattern was reversed. In the remaining family the mother spoke slightly more Dutch than English and the father always spoke Dutch. All but one child had siblings: eleven children were first-born, eight had one older sibling and one was the youngest of three. With four exceptions, all (older or same-age siblings) almost always spoke Dutch with the participating child. There were 12 children attending daycare, seven attending school and one child transitioning from daycare to school; the language of communication at all schools and daycares was Dutch. All participating families were high SES: with the exception of one father who had completed secondary education only, both parents had a university degree.

The United Kingdom sample was more heterogeneous and included families using the *one parent, one language* approach and families where both parents spoke the same language. More specifically, in five families the mother mostly spoke Dutch to the child and the father always spoke English, in one family this pattern was reversed, and in seven families both parents spoke Dutch to the child. In the remaining two families, both parents mostly (or always) spoke English. All but four children had siblings: four children were first born, six were the youngest with one or two older siblings, one was the middle child of three. The main language of communication amongst siblings was English. Thirteen children were exposed to English at nursery/preschool; one child stayed at home with an English-speaking childminder; and one child stayed at home with her English-speaking father. All participating families were high SES: with the exception of one father who had completed further education college, all parents held a university degree.

### Method and Procedure

To examine performance-based and experience-based measures of language dominance in bilingual English–Dutch, we used three sources of data: (1) children's spontaneous speech productions in naturalistic interactions with a parent or researcher in each language; (2) children's receptive vocabulary skills; and (3) parental questionnaire data.

### Spontaneous Speech Recordings

All children were video-recorded in a half-hour session in each language. In the Netherlands, each child was recorded interacting with the parent who normally used the language in question with the child. The parent was asked to interact with their child as they would usually do. Due to the heterogeneity of the United Kingdom sample and the unavailability of some of the English-speaking parents, children were usually recorded interacting with their parents (primarily mothers) in Dutch, whereas for English, all but one child were recorded interacting with a (near-)native-speaker research assistant. All children in

the Netherlands and all but two children in the United Kingdom were recorded in their homes. One was recorded at nursery rather than home because the child would only speak English at nursery; the other was recorded in the university's developmental lab at the parent's request. Irrespective of location or interlocutor, all children participated in similar activities in both languages, typically involving playing with puzzles or lego, looking at picture books or drawing.

The data were transcribed in CLAN/CHAT (MacWhinney, 2000) by a (near-)native-speaker of English or Dutch and checked for accuracy by another assistant. The following were excluded from analysis: incomplete utterances (e.g., trailing off), direct imitations of interlocutor, self-repetition, series of utterances (e.g., counting), utterances containing unintelligible parts, as well as any utterances in or containing words from the other language (with the exception of proper names and accepted loanwords).

In both samples and languages, we calculated the MLU in words as well as the average length of the longest five utterances in the sample (Upper Bound, UB5). The FREQ function was used to generate a list of words for each sample and the number of different verbs (VERBS) and nouns (NOUNS) was extracted and counted manually; any ambiguities were checked against the original transcript. Given the differences in sample size across children and languages and following common procedure in the field, data were analyzed for the first 100 utterances only; where fewer than 100 utterances were available, all utterances were included. For the children in the Netherlands, all but one produced at least 100 utterances in 30 min in Dutch. For English, 11 children did not reach 100 utterances. For the children in the United Kingdom, all produced 100 utterances in Dutch but four produced fewer than 100 utterances in English.

### Receptive Vocabulary Skills

In addition to the indicators of language abilities from children's spontaneous speech, we also assessed their receptive vocabulary skills, using standardized vocabulary tests. The PPVT-III-NL was used for Dutch (Dunn et al., 2005). For English, children in the Netherlands were given the PPVT-4 (Dunn and Dunn, 2007) or BPVS-2 for English, depending on the variety of English the child was exposed to; children in the United Kingdom completed the BPVS III (Dunn et al., 1997). Raw scores were converted to standard scores following the procedure in the manual; a score of between 85 and 115 indicates age-appropriate development for a monolingual child. The analyses rely on raw scores as standard scores were not available for children under the age of three. We used the raw scores to compare performance within and across children and languages, and to provide a general assessment of children's lexical knowledge.

### Parental Questionnaire

Information concerning the children's language experience was collected using an extensive parental questionnaire, the BiLEC (Bilingual Language Experience Calculator; Unsworth, 2013, following Gutiérrez-Clellen and Kreiter, 2003; Paradis, 2011). Parents were asked to indicate where and with whom the child spent time on an average day in the week and an average day at

the weekend, for how long, and which language(s) each person used when addressing the child, as well as time spent on extra-curricular activities and the language(s) in which these occurred. This information was used to calculate proportion of language exposure to Dutch vs. English at the current time (see Unsworth, 2013 for more details). Comparable information was gathered concerning the child's output with the same interlocutors and this was used to calculate current proportion of language use in Dutch vs. English at the current time. Finally, parents were asked about children's patterns of language exposure in the past and this was used to calculate their *cumulative* length of exposure (see Unsworth, 2013 for more details).

### Procedure

Children were tested on separate occasions in each language, with no more than 2 weeks between sessions, and for almost all children, the following test order was used: vocabulary task followed by spontaneous speech production. At the end of one of the two sessions, the parent completed the background and language experience questionnaire during a short informal interview.

## RESULTS

The experiential-based measures derived from the parental questionnaire are presented in **Table 1**. At the group level, children had a relatively balanced exposure to their two languages, but they used Dutch more frequently than English in both locations. There was considerable individual variation, suggesting a range of patterns of language exposure and use in the dataset.

The performance-based measures derived from the spontaneous speech samples and the vocabulary tests are given in **Table 2**. As noted in the Section "Materials and Methods," differential scores for vocabulary were calculated using the raw scores. The mean standard score for vocabulary was 100 ( $SD = 11.3$ ) for English and 99.5 ( $SD = 15.2$ ) for Dutch. At the group level, children tended to produce longer sentences in Dutch than in English, but the number of different nouns and verbs was more comparable across the two languages, as were the vocabulary scores. Once again, there was considerable variation between children, suggesting a range of patterns of language proficiency in the dataset.

Comparing children's scores in the two languages, MLU was significantly higher in Dutch than English ( $t(34) = 3.21, p = 0.003$ ) but there were no significant differences between languages on the other three scores (VERBS:  $t(34) = 0.600, p = 0.552$ ; NOUNS:  $t(34) = -0.653, p = 0.518$ ; UB5:  $t(34) = 0.550, p = 0.586$ ; VOCAB:  $t(34) = 1.63, p = 0.112$ ).

### Establishing the Strength and Shape of the Relationship Between Experience-Based and Performance-Based Measures

To explore the relationship between the experience-based and performance-based measures we first conducted bivariate

**TABLE 1** | Mean age, relative language exposure and use in Dutch, and cumulative length of exposure to Dutch and English ( $N = 35$ ).

	Age (months)	Average proportion of weekly input in Dutch (%)	Average proportion of weekly output in Dutch (%)	Cumulative length of exposure to Dutch (months)	Cumulative length of exposure to English (months)
Mean	43.2	53	59	24	18
SD	9.8	18	35	0.9	0.6

**TABLE 2** | Mean absolute and relative proficiency scores ( $N = 35$ ).

Variable	Absolute scores				Relative scores	
	Dutch		English		Differential (Dutch – English)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
MLU (in words)	2.99	0.92	2.44	1.00	0.54	0.99
UB5 (in words)	8.43	3.80	7.97	4.58	0.45	4.89
VERBS (#)	16.1	8.71	14.9	9.13	1.14	11.5
NOUNS (#)	17.0	8.62	18.1	9.07	–1.14	10.4
VOCAB (raw score)	50.9	18.0	46.5	19.6	4.40	15.9

correlational analyses to establish the strength of any relationships between the two sets of variables. Amount of exposure correlated significantly with differentials for MLU ( $r = 0.57$ ,  $p = 0.001$ ), UB5 ( $r = 0.57$ ,  $p = 0.001$ ), VERBS ( $r = 0.65$ ,  $p < 0.001$ ), NOUNS ( $r = 0.49$ ,  $p = 0.002$ ) and VOCAB ( $r = 0.49$ ,  $p = 0.003$ ) and language use correlated significantly with differentials for MLU ( $r = 0.77$ ,  $p < 0.001$ ), UB5 ( $r = 0.67$ ,  $p < 0.001$ ), VERBS ( $r = 0.84$ ,  $p < 0.001$ ), NOUNS ( $r = 0.55$ ,  $p = 0.001$ ) and VOCAB ( $r = 0.45$ ,  $p = 0.006$ ) (cf. original study, where there were only significant correlations with MLU<sub>diff</sub> and VERBS<sub>diff</sub>).

As a next step, linear ( $y = b_0 + b_1x$ ), quadratic ( $y = b_0 + b_1x + b_2x^2$ ) and cubic ( $y = b_0 + b_1x + b_2x^2 + b_3x^3$ ) relationships were estimated using the *Curve Estimation* function

in IBM SPSS v.25. The goal of this analysis was to determine whether the relation between the experience- and performance-based measures was best accounted for by non-linear rather than linear regression models (following Thordardottir, 2011 and Bedore et al., 2012). The results are presented in **Table 3**.

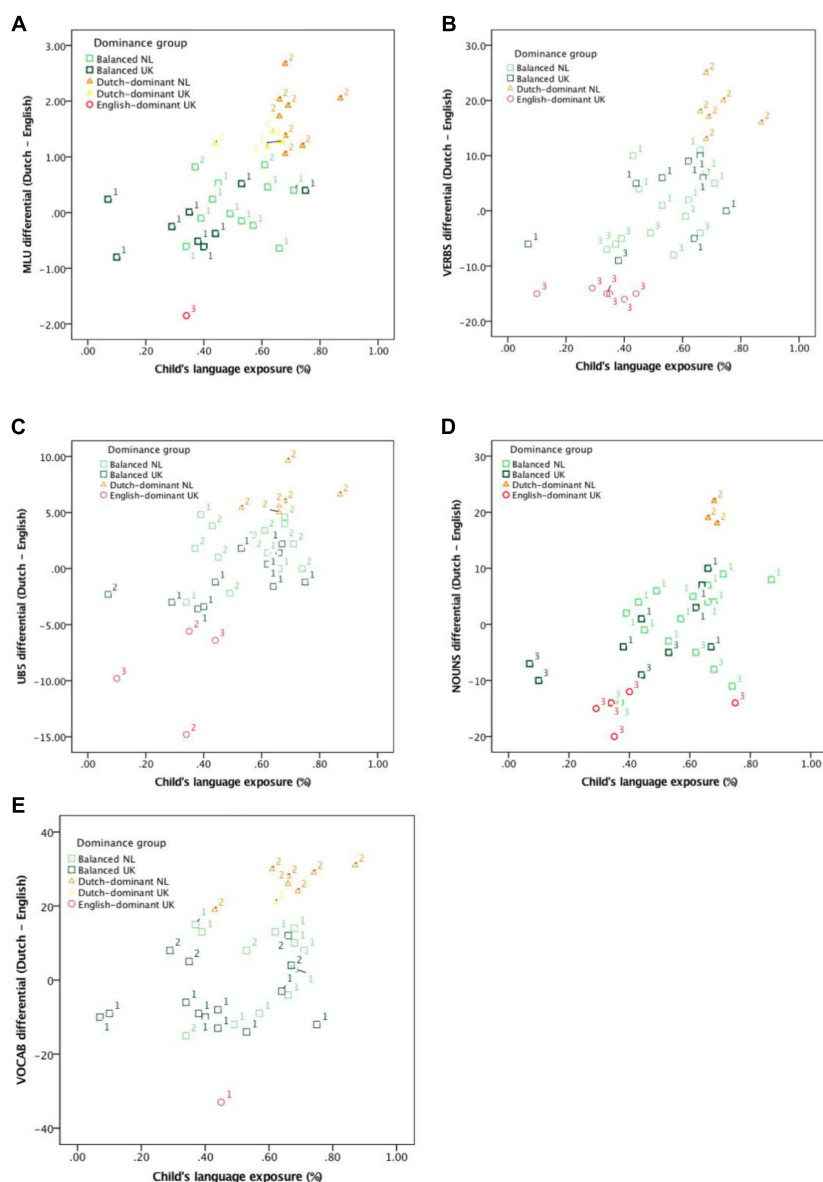
In all cases, the linear plus non-linear models accounted for more variance (i.e., had a higher total incremental  $R^2$  value) than the linear models alone, although the additional unique variance explained by the non-linear models was negligible in certain cases. The amount of variance explained for NOUNS<sub>diff</sub> was lower than for the other performance-based measures of relative proficiency, especially MLU<sub>diff</sub> and VERBS<sub>diff</sub>, and across most measures language use was a better predictor than language exposure.

**TABLE 3** | Summary of regression models using different estimation methods (incremental  $F$ -values and  $R^2$ ).

Variable	Type of regression model		
	Linear	Quadratic	Cubic
<b>MLU<sub>diff</sub></b>			
Exposure	$F = 23.9^{***}$ $R^2 = 0.42$	$F = 13.6^{***}$ $R^2 = 0.04$	$F = 9.7^{***}$ $R^2 = 0.02$
Use	$F = 50.9^{***}$ $R^2 = 0.61$	$F = 26.1^{***}$ $R^2 = 0.01$	$F = 17.2^{***}$ $R^2 = 0.01$
<b>UB5<sub>diff</sub></b>			
Exposure	$F = 21.6^{***}$ $R^2 = 0.40$	$F = 10.5^{***}$ $R^2 = 0.00$	$F = 7.3^{***}$ $R^2 = 0.02$
Use	$F = 27.0^{***}$ $R^2 = 0.45$	$F = 14.0^{***}$ $R^2 = 0.02$	$F = 9.4^{***}$ $R^2 = 0.01$
<b>VERBS<sub>diff</sub></b>			
Exposure	$F = 33.8^{***}$ $R^2 = 0.51$	$F = 18.3^{***}$ $R^2 = 0.03$	$F = 13.1^{***}$ $R^2 = 0.03$
Use	$F = 83.3^{***}$ $R^2 = 0.72$	$F = 42.0^{***}$ $R^2 = 0.01$	$F = 27.6^{***}$ $R^2 = 0.00$
<b>NOUNS<sub>diff</sub></b>			
Exposure	$F = 13.3^{***}$ $R^2 = 0.29$	$F = 6.5^{**}$ $R^2 = 0.00$	$F = 6.0^{**}$ $R^2 = 0.08$
Use	$F = 14.3^{***}$ $R^2 = 0.30$	$F = 7.1^{**}$ $R^2 = 0.01$	$F = 4.6^{**}$ $R^2 = 0.00$
<b>VOCAB<sub>diff</sub></b>			
Exposure	$F = 10.2^{**}$ $R^2 = 0.24$	$F = 5.8^{**}$ $R^2 = 0.03$	$F = 3.8^*$ $R^2 = 0.00$
Use	$F = 8.5^{**}$ $R^2 = 0.20$	$F = 4.3^*$ $R^2 = 0.01$	$F = 3.1^*$ $R^2 = 0.02$

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





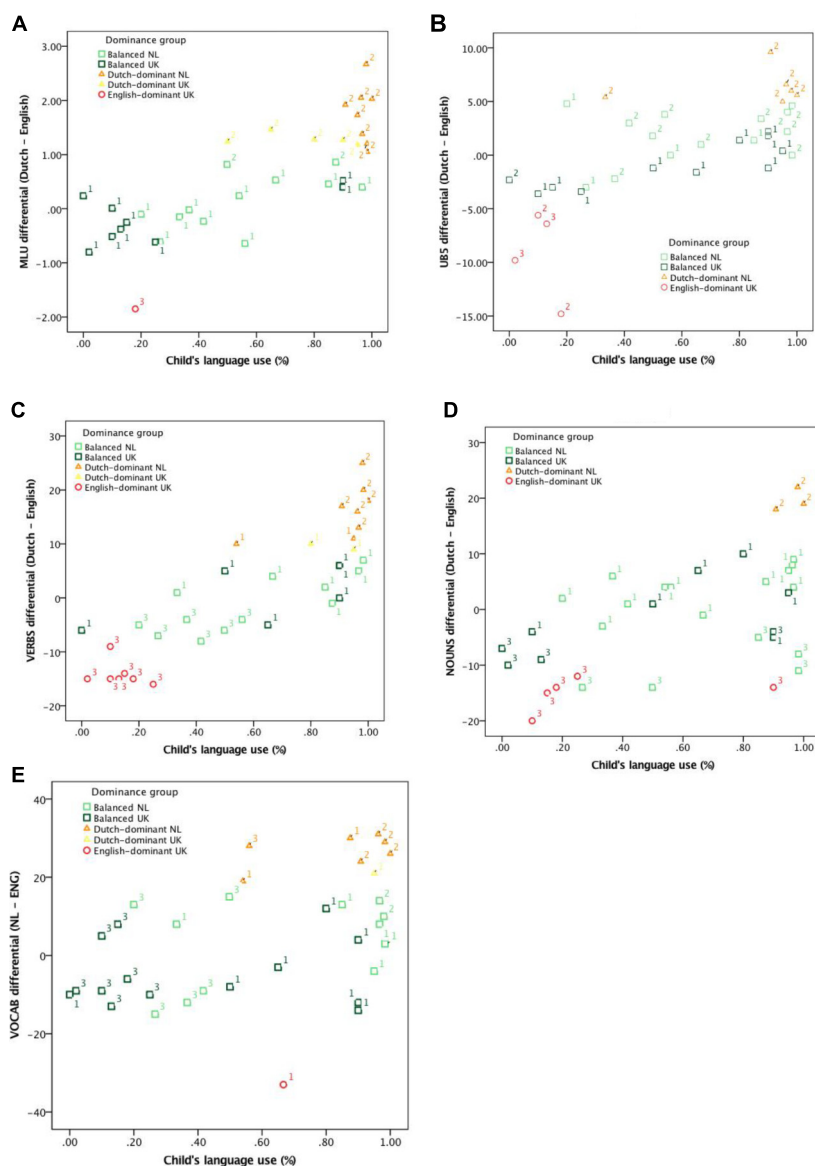
**FIGURE 1 |** Scatterplots showing relationship between relative language exposure (% Dutch) and relative language proficiency as measured by differentials for MLU (A), VERBS (B), UB5 (C), NOUNS (D), and VOCAB (E) for two different classifications dominance groups based on SDs and *k*-means cluster analysis groups.

1 = Balanced, 2 = Dutch-dominant, 3 = English-dominant according to the *k*-means cluster analysis.

## Predicting Dominance Group Membership Using Experiential Variables

Children were classified as Dutch-dominant, balanced, and English-dominant in exactly the same way as in the original study: children were classified as dominant in one of their two languages when there was a difference of greater than 1 SD between the two, and when this difference was less than 1 SD, children were classified as balanced. For example, children whose MLU in Dutch was greater than their MLU in English by at least 0.99 words (cf. **Table 2**) were considered Dutch-dominant, children whose MLU in English was greater than their MLU in Dutch by at least 0.99 words were

considered English-dominant and children with a differential score less than 0.99 words were classified as balanced. As noted in Unsworth (2016a), using a one-word difference in MLU as a measure for dominance is in line with earlier work (e.g., Bernardini and Schlyter, 2004). As in Unsworth (2016a), we extend this approach to the remaining performance-based variables using SDs as our guideline for dominance classification (see Discussion section for further consideration of this approach). The distribution of children in the three dominance groups, along with their country of residence, indicated by color in both cases, is presented in **Figure 1** in relation to language exposure and in **Figure 2** in relation to language use.



**FIGURE 2 |** Scatterplots showing relationship between relative language use (% Dutch) and relative language proficiency as measured by differentials for MLU (A), VERBS (B), UB5 (C), NOUNS (D), and VOCAB (E) for two different classifications dominance groups based on SDs and *k*-means cluster analysis groups. 1 = Balanced, 2 = Dutch-dominant, 3 = English-dominant according to the *k*-means cluster analysis.

Despite being resident in the United Kingdom, only a few of the children who were added to the original dataset as part of this study were English-dominant, and when children were classified into dominance groups on the basis of  $MLU_{diff}$ , there was just one.

A series of multinomial logistic regression analyses were run (in IBM SPSS v.25) to investigate how well language exposure and use predicted group membership (balanced, Dutch-dominant, English-dominant) for the dominance groups derived on the basis of the four different performance-based measures ( $MLU_{diff}$ ,  $VERBS_{diff}$ ,  $NOUNS_{diff}$  and  $UB5_{diff}$ ). The reference category was set to English-dominant. The extent to which language use and exposure correctly predicted group membership, the

overall success rate of the model in doing so, as well as an estimation of the amount of variance explained (Nagelkerke's *R*) are given in **Table 4**. Values between 80 and 89% are considered fair, while rates above 90% are good (Plante and Vance, 1994).

For  $MLU_{diff}$ , language exposure was a good predictor for group membership; the model with language exposure as a predictor against a constant-only model was statistically significant ( $\chi^2 = 17.55$ ,  $p < 0.001$ ,  $df = 2$ ), as was the model with language use as a predictor ( $\chi^2 = 19.64$ ,  $p < 0.001$ ,  $df = 2$ ). Prediction success overall was high for language exposure and fair for language use. In both cases, the model was better at predicting group membership for the Dutch-dominant children than the

balanced children and prediction success was poor for the single English-dominant child.

For UB5<sub>diff</sub>, language exposure and language use were equally good predictors for group membership and both models with the two predictors against a constant-only model were statistically significant (exposure:  $\chi^2 = 12.3$ ,  $p = 0.002$ ,  $df = 2$ ; use:  $\chi^2 = 15.5$ ,  $p < 0.001$ ,  $df = 2$ ). The model was better at predicting group membership for the balanced children than for the other two groups for both language exposure and use.

For VERBS<sub>diff</sub>, language use was a better predictor than language exposure for groups membership although both predictors had good prediction success (exposure:  $\chi^2 = 21.6$ ,  $p < 0.001$ ,  $df = 2$ ; use:  $\chi^2 = 30.3$ ,  $p < 0.001$ ,  $df = 2$ ). The model with language exposure was better at classifying the balanced children and poor at classifying the other two groups, whereas the model with language use was good at classifying both balanced and Dutch-dominant children.

For NOUNS<sub>diff</sub>, language exposure was not a good predictor for group membership and the model with exposure as a predictor against a constant-only model was not statistically significant ( $\chi^2 = 1.5$ ,  $p = 0.22$ ,  $df = 1$ ). Overall prediction success was fair. However, the model could only accurately classify balanced children. The results were identical for language use ( $\chi^2 = 4.45$ ,  $p < 0.05$ ,  $df = 1$ ).

For VOCAB<sub>diff</sub>, the models with language exposure and language use as predictors differed marginally from the constant only model and their overall prediction accuracy was rather poor (exposure:  $\chi^2 = 7$ ,  $p = 0.03$ ,  $df = 2$ , Nagelkerke = 0.25; use:  $\chi^2 = 6.8$ ,  $p = 0.034$ ,  $df = 2$ , Nagelkerke = 0.24). Both models had excellent accuracy at predicting groups membership for the Dutch-dominant group, but otherwise, their prediction accuracy was poor for the other two groups.

In each of these analyses, the number of cases per level of the dependent measure (i.e., the number of children per group) varied considerably and in some cases (e.g., English-dominant children in the analysis for MLU<sub>diff</sub>), the number of cases was extremely low. This means that the analysis may be biased and there may be complete separation in the data (King and Zeng, 2001). To address this potential problem, we ran the analysis using penalized maximum likelihood estimation (Firth, 1993). We used the logistf package (Heinze et al., 2016) in R (R Development Core Team, 2008) to run a penalized regression analysis. The results were comparable in terms of the prediction success, with the only difference for UB5<sub>diff</sub> only: with respect to language exposure, group membership was predicted correctly for 3 out of 6 (50%) Dutch-dominant children, 18 out of 25 (72%) balanced children and 1 out of 4 (25%) English-dominant children, and with respect to language use, the values were 0% (0/6), 96% (24/25) and 25% (1/4), respectively. However, only the models for MLU<sub>diff</sub> and VOCAB<sub>diff</sub> were significant. Furthermore, the amount of variance differed across analyses with (in some cases) substantially lower values for the  $R^2$  for the penalized model than for the Nagelkerke's  $R$  calculated for the standard model (compare the two rightmost columns in Table 4).

To summarize, language exposure and use were good predictors of group membership when this was based on

differentials using MLU<sub>diff</sub>, VERBS<sub>diff</sub> and UB5<sub>diff</sub>, but this did not hold for differentials using NOUNS<sub>diff</sub> and VOCAB<sub>diff</sub>. When taking into account the small sample size, the statistical models accounted for less variance, but the overall patterns observed were comparable.

## Independently Verifying Dominance Groups and Their Relation With Language Exposure and Use

To investigate whether the dominance groups derived using the standard deviations on the various performance-based measures (differential scores) were valid, we ran two different types of cluster analysis (following Cattani et al., 2014), namely a hierarchical agglomerative cluster (HAC) analysis and a  $k$ -means cluster analysis using Ward's minimum variance method (Ward, 1963) and squared Euclidean distance as the similarity measure. HAC is a bottom-up method used to determine the number of clusters in the dataset without predetermining the possible number of clusters. In a  $k$ -means cluster analysis, the number of (potential) clusters is pre-specified by the researcher. For both methods, we used the R package NbClust package (Charrad et al., 2014) to independently determine the optimal number of clusters. This package allows the researcher to simultaneously run up to 30 different indices, including the commonly used Gap statistic and silhouette index; the optimal number of clusters reported here is the value given by the majority of indices. We compared the clusters resulting from the HAC and  $k$ -means analyses with our own classification and examined the extent to which these overlapped. Subsequently, we investigated the relation between the clusters resulting from the  $k$ -means analysis and language exposure and use, and once again compared this to our original classification.

For both analyses, children's scores on each measure were entered into the models without any information about either group membership (i.e., Dutch- or English-dominant or balanced) or language exposure or use. Standardly, Hopkins statistic (Hopkins and Skellam, 1954) can be used as an indication of the clusterability of a dataset. Values above 0.5 are typically interpreted as evidence that the data in a given sample are not uniformly distributed; in other words, values under 0.5 suggest that the data may not be clusterable. However, as noted by Banjaree and Dave (2004), applying Hopkins statistic to small datasets is problematic, if not impossible. We report the values here for the sake of completeness but with this caveat ( $H = 0.395$  for MLU<sub>diff</sub>, 0.415 for UB5<sub>diff</sub>, 0.625 for VERBS<sub>diff</sub>, 0.604 for NOUNS<sub>diff</sub> and 0.499 for VOCAB<sub>diff</sub>). For three out of the five performance-based measures – VERBS<sub>diff</sub>, UB5<sub>diff</sub>, and VOCAB<sub>diff</sub> – the number of clusters generated by the HAC analysis was the same as in our classification (i.e., three). For NOUNS<sub>diff</sub>, the HAC analysis generated four rather than three clusters, and for MLU<sub>diff</sub>, the optimal number of clusters was between three and five, depending on the index used.

To explore the composition of the various clustering options for MLU<sub>diff</sub>, we generated dendrograms for the clusters proposed by the HCA. A comparison of the dendrograms for the analysis for MLU<sub>diff</sub> with the least (three) and most (five)

**TABLE 4 |** Estimates of language exposure and use on dominance group membership based on differential scores (% and number of children correctly predicted).

	Prediction success				Variance explained	
	Overall	NL-dom	Balanced	ENG-dom	Nagelkerke's R (standard regression model)	R <sup>2</sup> (Firth penalized regression model)
<b>MLU<sub>diff</sub></b>						
Exposure	83%	92%(12/13)	81% (17/21)	0% (0/1)	0.50***	0.32**
Use	77%	85% (11/13)	76% (16/21)	0% (0/1)	0.55***	0.35***
<b>UB5<sub>diff</sub></b>						
Exposure	74%	17% (1/6)	96% (24/25)	25% (1/4)	0.37***	0.006 <sup>n.s.</sup>
Use	71%	0% (0/6)	96% (24/25)	25% (1/4)	0.45***	0.034 <sup>n.s.</sup>
<b>VERBS<sub>diff</sub></b>						
Exposure	66%	33% (2/6)	87% (20/23)	17% (1/6)	0.56***	0.05 <sup>n.s.</sup>
Use	80%	83% (5/6)	83% (19/23)	67% (4/6)	0.70***	0.07 <sup>n.s.</sup>
<b>NOUNS<sub>diff</sub></b>						
Exposure	77%	0% (0/3)	100% (27/27)	0% (0/5)	0.15 <sup>n.s.</sup>	0.06 <sup>n.s.</sup>
Use	77%	0% (0/3)	100% (27/27)	0% (0/5)	0.33***	0.06 <sup>n.s.</sup>
<b>VOCAB<sub>diff</sub></b>						
Exposure	74%	13% (1/8)	92% (25/26)	0% (0/1)	0.25*	0.15*
Use	74%	0% (0/8)	100% (26/26)	0% (0/1)	0.24*	0.21*

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

**TABLE 5 |** Number of children assigned to each cluster ( $k$ -means cluster analysis), their distribution across language exposure/use groups and mean scores per measure.

Measure	Cluster	N	Language exposure			Language use			Mean score (range)	Comparison across clusters (ANOVA)
			≤35%	<65%	≥65%	≤10%	<90%	≥90%		
MLU	Cluster 1 (≈ Balanced)	19	4	12	3	3	13	3	0.86 (0.24 – 1.46)	$F(2,34) = 68, p < 0.001$
	Cluster 2 (≈ Dutch-dominant)	15	0	5	10	0	9	10	2.08 (1.73 – 2.67)	
	Cluster 3 (≈ English-dominant)	1	1	0	0	1	0	0	1.85	
UB5	Cluster 1 (≈ Balanced)	14	2	7	5	1	10	3	−2.36 (−6.40 – 0.40)	$F(2,34) = 53.3, p < 0.001$
	Cluster 2 (≈ Dutch-dominant)	19	3	5	11	2	7	10	3.87 (1 – 9.60)	
	Cluster 3 (≈ English-dominant)	2	1	1	0	1	1	0	−12.36 (−14.8 – 9.8)	
VERBS	Cluster 1 (≈ Balanced)	16	1	8	7	1	10	5	4 (−6 – 11)	$F(2,34) = 100.2, p < 0.001$
	Cluster 2 (≈ Dutch-dominant)	6	0	4	2	0	4	2	18.17 (13 – 25)	
	Cluster 3 (≈ English-dominant)	13	5	4	4	3	6	4	−10.23 (−16 – −4)	
NOUNS	Cluster 1 (≈ Balanced)	18	2	5	11	1	11	6	3.28 (−4 – 10)	$F(2,34) = 68.8, p < 0.001$
	Cluster 2 (≈ Dutch-dominant)	3	0	1	2	0	0	3	19.67 (18 – 22)	
	Cluster 3 (≈ English-dominant)	14	3	7	4	3	8	3	−11.29 (−20 – −5)	
VOCAB	Cluster 1 (≈ Balanced)	15	1	8	6	1	6	8	16.15 (3 – 15)	$F(2,34) = 100.5, p < 0.001$
	Cluster 2 (≈ Dutch-dominant)	6	0	0	6	0	0	6	−11.13 (19 – 31)	
	Cluster 2 (≈ English-dominant)	12	5	6	1	3	9	0	−11.13 (−33 – −3)	

clusters generated revealed two differences. First, the three-cluster classification generated a single group at the lower end with children scoring between −1.85 and 0.10, whereas the five-cluster classification created two subgroups at the same lower end, separating one child with a score of −1.85 from the rest of the group. Second, the five-cluster classification identified an extra subgroup in the middle range of the distribution, separating children with scores between 0.24 and 0.53 from children with scores between 0.82 and 1.46, whereas in the three-cluster dendrogram these two groups were collapsed into a single group. In other words, whilst by definition

more fine-grained, the partitioning of the children provided by the five-cluster classification was qualitatively comparable with the broader three-cluster classification. For this reason, we decided to adopt a three-cluster grouping in the  $k$ -means analysis (i.e., where the number of clusters is pre-specified by the researcher) in order to maximize the comparability of this independent means of classifying children and our original classification; given our relatively small sample size, the three-cluster option would also maximize the number of children in each cluster and for this reason was also preferable.



We adopted the same approach for NOUNS<sub>diff</sub> and subsequently compared the dendrogram for the bottom-up four-cluster solution with the top-down three-way classification in order to establish their comparability. The two different cluster analyses overlapped at the upper end of the distribution (the same three participants with scores between 18 and 22 were in one cluster on both analyses). The only difference was at the lower end of the distribution: whereas the three-cluster classification generated a single group with children scoring between  $-20$  and  $-5$ , the four-cluster classification divided these participants across two subgroups separating children with scores between  $-20$  and  $-14$  from children with scores between  $-12$  to  $-3$ . In other words, at the broader level the three- and four-cluster solutions divide the group in qualitatively comparable places.

**Table 5** presents the results of the *k*-means cluster analysis, where we set the number of clusters at three. The goal of this analysis was to determine whether children were grouped in a similar way as our own classification. As such, each of the three clusters is labeled as English-dominant, Dutch-dominant or balanced, depending on the distribution of the children's scores within that cluster. In addition, in order to compare the relationship between clusters and experiential variables across the cluster analysis and our own classification, **Table 5** specifies the number of children in each cluster with more than 65% exposure to Dutch (i.e., children who are expected to be Dutch-dominant), children with less than 35% exposure to Dutch (i.e., children who are expected to be English-dominant), as well as those who fall in between (and are thus expected to be balanced); similarly, it also specifies the number of children who use Dutch at least 90% of the time (i.e., children who are expected to be Dutch-dominant), children who use Dutch no more than 10% of the time (i.e., children who are expected to be English-dominant) and those who fall in between (and are thus expected to be balanced).

For all measures except VOCAB<sub>diff</sub>, children are distributed over the different language exposure/use groups more or less as expected, with most of the children in the "balanced" cluster falling in the mid-range ( $<65\%$  exposure to Dutch,  $<90\%$  use of Dutch), most in the "Dutch-dominant" cluster in the highest range ( $\geq 65\%$  exposure,  $\geq 90\%$  use) and most in the "English-dominant" cluster in the lowest range ( $\leq 35\%$  exposure,  $\leq 10\%$  use). It should be noted, however, that there are almost always exceptions, and the relative distribution of children across the various clusters for NOUNS<sub>diff</sub> and for the "English-dominant" cluster for VERBS<sub>diff</sub> and VOCAB<sub>diff</sub> is different; in the latter case, the "cut-off" point for dominance in English appears to lie around the 30% mark rather than 10% (cf. **Figure 2**).

As a last step, we visually examined the relationship between language exposure (**Figure 1**) and language use (**Figure 2**) with our two types of classification, that is, the original SD-based classification referred to in the Figures as *Dominance group* and represented using colors, and the *k*-means cluster analysis, represented with numbers (cf. **Table 5**).

**Figures 1, 2** reveal that these two ways of classifying children overlap considerably. In the case of MLU<sub>diff</sub>, the children classified as Dutch- or English-dominant on our analysis were also grouped together in the *k*-means cluster analysis and the

same holds for all but two of the 21 balanced children. A similar pattern holds for VERBS<sub>diff</sub> and NOUNS<sub>diff</sub>, with all Dutch- or English-dominant children together in the same groups on the *k*-means cluster analysis as in our classification, and the same for 16 of the 23 balanced children for VERBS<sub>diff</sub> and 18 of the 27 balanced children for NOUNS<sub>diff</sub>. For VOCAB<sub>diff</sub> and UB5<sub>diff</sub> the two means of grouping children differ slightly more: for VOCAB<sub>diff</sub>, 4 of the 7 Dutch dominant children, neither the two English-dominant children and 15 of the balanced children on the SD-based classification are grouped together in the *k*-means cluster analysis. For UB5<sub>diff</sub> there is overlap for almost half the children the SD and the *k*-means cluster classification. In short, then, the main difference between the *k*-means cluster analysis and our classification is that the cluster analysis grouped fewer children together in the middle of the distribution; in most cases (VERBS<sub>diff</sub>, NOUNS<sub>diff</sub> and VOCAB<sub>diff</sub>), the cluster analysis grouped the children at the lower end together with the children classified as English-dominant (i.e., in red) in the original analysis.

To summarize, the hierarchical and the *k*-means cluster analyses grouped the children into largely comparable groups as the performance-based classification, and by and large the relationship between these groups, on the one hand, and language exposure and use, on the other, were also similar.

## DISCUSSION

In this paper we examined the relationship between experiential-based and performance-based measures of language dominance in bilingual English-Dutch children in the United Kingdom and the Netherlands. More specifically, using parental questionnaire data we derived estimates of children's patterns of language exposure and use and related these to differential scores for five variables derived from spontaneous speech data, namely morphosyntactic complexity measured by the MLU and the mean length of the longest five utterances (UB5), and lexical diversity measured by the number of different verb types (VERBS), the number of different noun types (NOUNS) and scores on a standardized vocabulary task.

### The Relationship Between Relative Exposure and Use and Relative Proficiency

Our first research question asked to what extent experiential-based factors were related to bilingual children's relative language proficiency. The findings revealed a moderate to strong relationship between relative exposure and relative use, on the one hand, and relative proficiency as measured by MLU<sub>diff</sub>, VERBS<sub>diff</sub>, UB5<sub>diff</sub>, NOUNS<sub>diff</sub> and VOCAB<sub>diff</sub>, on the other. The observation that such a relationship exists for all five outcome variables is in contrast to the original study (Unsworth, 2016a), where only MLU<sub>diff</sub> and VERBS<sub>diff</sub> were found to have a significant relation with language exposure and use. This difference is most likely the result of a larger sample and/or including children from across the dominance continuum. For example, the number of children with more exposure to English

than Dutch more than doubled from six in the original sample to fifteen with the inclusion of the United Kingdom-resident children.

Curve-fitting analyses revealed that, as predicted, the relationship between relative experience and relative proficiency was generally best accounted for with a non-linear model. This is in line with previous research exploring differential effects of exposure and use on absolute measures of proficiency (e.g., Thordardottir, 2011; Cattani et al., 2014) and with Bedore et al. (2012) larger-scale study. It should be noted, however, that whilst more variance was captured by the non-linear models, the additional unique variance which they explain was limited (between 1 and 8%) and negligible (<1%) in certain cases.

### Predicting Relative Proficiency Using Relative Exposure and Use, and Vice Versa

Our second research question asked whether language exposure and use could reliably classify bilingual children into language dominance groups. To this end, children were initially classified into Dutch-dominant, English-dominant and balanced groups using the standard deviation for the variable in question as cut-off point. For the number of different verb types, for example, this meant that children in the English-dominant group produced more than 12 different verb types in English than in Dutch; for the Dutch-dominant group, this pattern was reversed and for the balanced group the difference in number of verb types across the two languages was no more than 12.

By and large, the children in this larger sample patterned as predicted on the basis of the smaller sample in the original study (Unsworth, 2016a). On the whole, language exposure and use were best at predicting group membership when this was based on  $MLU_{diff}$ , with accurate classification for around four fifths of the children. A similar pattern was observed for language use as a predictor of group membership based on  $VERBS_{diff}$ . Classification was less accurate, but still greater than 50%, for language exposure and  $VERBS_{diff}$  and for both experiential variables as predictors of group membership based on  $UB5_{diff}$ , whereas for  $NOUNS_{diff}$  and  $VOCAB_{diff}$ , classification was poor. The amount of variance accounted for and the significance of the model depended on the analysis, with the clearest results for  $MLU_{diff}$  and  $VOCAB_{diff}$ .

When derived independently, rather than using the somewhat arbitrary standard deviation as cut-off point, a three-cluster solution was the optimal analysis (or one of the optimal analyses) for four of the five measures (i.e.,  $MLU_{diff}$ ,  $VERBS_{diff}$ ,  $UB5_{diff}$ , and  $VOCAB_{diff}$ ); for  $MLU_{diff}$  four- and five-cluster solutions were also considered optimal and for  $NOUNS_{diff}$  there were four clusters. In order to maximize comparability with our own classification and because the analyses with more than three clusters were at a broader level qualitatively parallel with a three-way grouping, the *k*-means cluster analysis was pre-specified at three groups. For each of these three groups, the children's patterns of language exposure and use largely corresponded with the patterns observed in our first analysis. For example, most of the children who fell in the Dutch-dominant cluster had more than 65% exposure to Dutch and used Dutch for at least 90% of the time, at least for  $MLU_{diff}$ ,  $VOCAB_{diff}$ , and

$NOUNS_{diff}$ . In short, then, the results of the present study suggest that when relative proficiency is operationalised in terms of differentials, relative language exposure and language use can be used to classify children into dominance groups with a reasonable degree of success, and this especially holds for morphosyntactic proficiency.

### Language Exposure vs. Language Use

Children's relative proficiency scores were related to two aspects of their language experience, namely language exposure and language use. In general, the curve-fitting analyses revealed a stronger relation between relative language proficiency and relative language use than between relative language proficiency and relative language exposure. The point at which (the majority of) children were classified as Dutch-dominant, as opposed to balanced, also differed for language exposure (around 65%) and language use (around 90%): children who were classified as Dutch-dominant used more Dutch than they were exposed to and the same pattern held for the English-dominant children, too. Note, however, that for English-dominant children, the proportion of their language use in that language was less (around the 70% mark) than the equivalent value for Dutch-dominant children in Dutch (around 90%). In short, then, children who were classified as dominant in one of their two languages were those who used one of their two languages more than they heard it.

These findings align well with a recent study on similar-aged Spanish-English bilingual children by Ribot et al. (2017). These authors found that language use at 30 months predicted rate of acquisition in English for expressive skills, as measured by a single-word picture-naming task, but not for receptive skills, as measured by a more comprehensive language proficiency task targeting various aspects of semantics, morphology, syntax and preliteracy skills. More specifically, children whose use of English at 30 months was greater than their exposure to English had better picture-naming skills in English than children for whom exposure was greater than use. Notwithstanding the fact that the tasks used to assess abilities in the two modalities were not entirely comparable, Ribot et al. (2017, p. 8) speculated that this finding might be explained in two different ways: it may reflect a more general effect of language use which was observed only for expressive skills because these are harder to achieve, or it might constitute a specific effect whereby language use specifically benefits expressive skills. In the present study, relative language use was more weakly associated with receptive than expressive skills; for some of the analyses, however, a significant relationship was observed nevertheless. In principle then, the two explanations put forward by Ribot et al. (2017) could apply here, too. The most important parallel between these two studies is the following: what seems to be crucial is not the amount of language use *per se*, but the discrepancy between language use and language exposure, that is, children who were found to be dominant in one of their two languages tended to use that language more than they heard it.

In the interests of transparency, it is worth noting that the way in which language exposure and use were calculated in the present study was not completely equivalent: the language

use variable is based on the child's language use within the home only, whereas the language exposure variable also includes sources outside the home. This means that the children resident in the Netherlands likely used (even) more Dutch than estimated here and the children resident in the United Kingdom probably used less Dutch. Incorporating these differences into our analysis would mean that the United Kingdom-resident children should shift leftward in **Figure 2**, whereas the children resident in the Netherlands would shift rightward. If anything, this would only serve to make the cut-off point between balanced and Dutch-dominant children for language use more extreme.

## Measures of Morphosyntactic vs. Lexical Proficiency

The present study applied a number of performance-based measures to spontaneous speech samples. Following previous research on language dominance in early child bilinguals (e.g., Cantone et al., 2008), morphosyntactic complexity was assessed using MLU and upper bound and children's lexical diversity was measured using number of different noun and verb types and scores on a standardized receptive vocabulary test. On the whole, the relation between experience-based measures and performance-based measures of language dominance was clearest for MLU at the level of morphosyntax, and for number of different verb types at the lexical level. It is possible that VERBS may in part reflect morphosyntactic complexity in the sense that producing a range of verb types may in part reflect more complex, multi-verb utterances (see Unsworth, 2005, Chapter 4 for relevant discussion) and in this sense, be indicative of more complex grammatical structure. The use of nouns may be less reliable as an indicator of children's expressive skills: children may use various referring expressions – pronouns and demonstratives or null arguments – in some contexts to represent the subject and object arguments of their verbal utterances, but none of these are included in the noun count. They may produce morpho-syntactically complex sentences, and at the same time have comparatively lower scores on lexical diversity as estimated by noun type. This may underestimate their lexical abilities and potentially lead to less variation between children for this variable. It is then perhaps unsurprising that the dominance groups based on the number of different noun types were less differentiated and bore limited (if any) relation with experiential factors.

## Coarse vs. More Fine-Grained Measures of Language Dominance

The present study highlights important differences between coarser and more fine-grained measures of language exposure. For some children there appears to be a general effect of the language of the environment, that is, there were children resident in the United Kingdom who on the basis of relative exposure (parental questionnaire data) would be expected to fall within the balanced group but who were in fact stronger in English than Dutch (cf. **Figure 1**). Similarly, the sample included children resident in the Netherlands who on the basis of relative exposure

estimates were expected to fall within the balanced group but who were in fact stronger in Dutch than English (cf. **Figure 1**). It is likely that this reflects a more general effect of the language environment not captured by even the most detailed language background questionnaire (Pearson, 2007). Conversely, there were children in the Netherlands who were not Dutch-dominant and likewise, there were children resident in the United Kingdom who were not English-dominant. This was most likely because almost half of the families in the United Kingdom sample adopted a *minority language at home* approach, but there were also United Kingdom-resident children in *one parent, one language* families who were not English-dominant.

Taken together, these findings suggest that even though it is often used as such (e.g., Argyri and Sorace, 2007; Foroodi-Nejad and Paradis, 2009; Serratrice et al., 2009), language of the environment is not an accurate proxy for language dominance. This observation is consonant with recent findings by Hervé et al. (2016) and Schmeißer et al. (2016); in this latter study, the authors showed that language exposure at the individual level was a better predictor of the magnitude of cross-linguistic influence in bilingual children's language production than language exposure at the group level (i.e., country of residence).

## Limitations and Future Research

There are a number of limitations to the present study. First, the sample size remains relatively small and the family language constellations in the United Kingdom-resident children is more varied than in the Netherlands-based children. Second, some of the performance-based measures used here, in particular MLU, may not be amenable to cross-linguistic comparison for certain languages and/or language pairs (Yip and Matthews, 2006; Allen and Dench, 2015). Third, the present study focuses on differentials as a measure of language dominance. An alternative approach would be to calculate the between-languages ratio, that is, dividing a child's score for one language with his or her score for the other language (e.g., Sheng et al., 2014; Goriot et al., 2018) or to combine the two (Birdsong, 2016). Finally, with perhaps the exception of MLU, the values used to divide children into dominance groups are in a certain sense arbitrary in that they are sample-specific. To further assess the validity of these values for other samples, as well as the generalisability of the approach put forward here as a whole, future research should investigate different language combinations for different age groups and with different outcome measures.

## CONCLUSION

By using language proficiency measures commonly adopted in much of the previous literature on dominance in bilingual first language acquisition (e.g., Cantone et al., 2008) and relating these to experiential measures frequently used in the burgeoning literature on input effects, the present study brings together these two different strands of research in the field. In doing so, it expands Bedore et al. (2012) findings to younger bilingual children (i.e., 2- to 4-year-old children cf. 5 year olds) and to a different language combination (i.e., English–Dutch instead of



English–Spanish). Furthermore, this study shows that relative amount of exposure and relative amount of use can be used as a proxy for language dominance, understood in its narrow sense of relative language proficiency. Crucially, however, the relation between relative language proficiency and language experience differs for these two variables. It is exactly this difference which may make it possible to distinguish between dominant and balanced children. Given that measures such as language exposure and use more readily allow for cross-study comparisons than measures which are specific to certain age ranges, languages or studies (Grosjean, 1998), and whether a child produces more than she hears is in principle relatively easy to establish, this is a welcome finding.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of PPLS Research Ethics Guidelines, PPLS Ethics Committee. The protocol was approved by the PPLS Research Ethics Committee, University of Edinburgh. We obtained written informed consent from all parents/guardians

## REFERENCES

- Allen, S. E. M., and Dench, C. (2015). Calculating mean length of utterance for eastern Canadian Inuktitut. *First Lang.* 35, 377–406. doi: 10.1177/0142723715596648
- Argyri, E., and Sorace, A. (2007). Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism* 10, 79–99. doi: 10.1017/S1366728906002835
- Banjaree, A., and Dave, R. N. (2004). “Validating clusters using the Hopkins statistic,” in *Proceedings of the IEEE International Conference on Fuzzy Systems*, (Budapest: IEEE), 149–153. doi: 10.1109/FUZZY.2004.1375706
- Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., et al. (2012). The measure matters: language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism* 15, 616–629. doi: 10.1017/S1366728912000090
- 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
- Birdsong, D. (2016). “Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness,” in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 85–105. doi: 10.1017/CBO9781107375345.005
- Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *Int. J. Biling.* 14, 422–446. doi: 10.1177/1367006910370917
- Bohman, T. M., Bedore, L. M., Pena, E. D., Mendez-Perez, A., and Gillam, R. B. (2010). What you hear and what you say: language performance in Spanish–English bilinguals. *Int. J. Biling. Educ. Biling.* 13, 325–344. doi: 10.1080/13670050903342019
- Cantone, K. F., Müller, N., Schmitz, K., and Kupisch, T. (2008). Rethinking language dominance. *Linguist. Ber.* 215, 307–343.
- 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? *Int. J. Lang. Commun. Disord.* 49, 649–671. doi: 10.1111/1460-6984.12082
- Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36. doi: 10.18637/jss.v061.i06
- Chondrogianni, V., and Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguist. Approaches Biling.* 1, 318–345. doi: 10.1075/lab.1.3.05cho
- Döpke, S. (1992). *One Parent, One Language: An Interactional Approach*. Amsterdam: John Benjamins. doi: 10.1075/sibil.3
- Dunn, L. M., and Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (PPVT-4)*. Minneapolis, MN: Pearson.
- Dunn, L. M., Dunn, L. M., and Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Pearson.
- Dunn, L. M., Dunn, L. M., Whetton, C., and Burley, J. (1997). *The British Picture Vocabulary Scale*. London: GL Assessment.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27–38. doi: 10.1093/biomet/80.1.27
- Foroodi-Nejad, F., and Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in Persian–English bilinguals. *Bilingualism* 12, 411–427. doi: 10.1017/S1366728909990241
- 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
- Goriot, C., van Hout, R., Unsworth, S., McQueen, J., and Broersma, M. (2018). Language balance and switching ability in children acquiring English as a second language. *J. Exp. Child Psychol.* 173, 168–186. doi: 10.1016/j.jecp.2018.03.019
- Grosjean, F. (1982). *Life With Two Languages: An Introduction to Bilingualism*. Cambridge, MA: Harvard University Press.
- Grosjean, F. (2010). *Bilingual. Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459
- Grosjean, F. (1998). *Studying Bilinguals: Methodological and Conceptual Issues*. Bilingualism: Language and Cognition, 1, 131–149. doi: 10.1017/S136672899800025X
- Grüter, T., Hurtado, N., Marchman, V. A., and Fernald, A. (2014). “Language exposure and online processing efficiency in bilingual development: relative versus absolute measures,” in *Input and Experience in Bilingual Development*, eds T. Grüter and J. Paradis (Amsterdam: John Benjamins), 15–36. doi: 10.1017/CBO9781107415324.004
- Gutiérrez-Clellen, V. F., and Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Appl. Psycholinguist.* 24, 267–288. doi: 10.1017/S0142716403000158

of the children who participated in the study in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

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

## FUNDING

The present study was partly funded by the School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

## ACKNOWLEDGMENTS

Thanks are due to all participating families for their cooperation, to Patty Ernst for collecting the Dutch data in the UK, Loes Burgers and Chantal van Dijk for collecting and transcribing data in the Netherlands, and to Dafne van Leeuwen and Liz Smeets for further assistance.



- Hammer, C. S., Komaroff, E., Rodriguez, B. L., Lopez, L. M., Scarpino, S. E., and Goldstein, B. (2012). Predicting Spanish – English bilingual children's language abilities. *J. Speech Lang. Hear. Res.* 55, 1251–1264. doi: 10.1044/1092-4388(2012/11-0016)
- Heinze, G., Ploner, M., Dunkler, D., and Southworth, H. (2016). *Package 'logistf'*. Available at: <http://cemsis.meduniwien.ac.at/en/kb/science-research/software/statistical-software/fllogistf/>
- Hervé, C., Serratrice, L., and Corley, M. (2016). Dislocations in French–English bilingual children: an elicitation study. *Bilingualism* 19, 987–1000. doi: 10.1017/S1366728915000401
- 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
- Hopkins, B., and Skellam, J. G. (1954). A new method of determining the type distribution of plant individuals. *Ann. Bot.* 18, 213–227. doi: 10.1093/oxfordjournals.aob.a083391
- King, G., and Zeng, L. (2001). Logistic regression in rare events data. *Polit. Anal.* 9, 137–163. doi: 10.1093/oxfordjournals.pan.a004868
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*, 3rd Edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism*. Amsterdam: John Benjamins. doi: 10.1075/sibil.39
- Montrul, S. (2016). "Dominance and proficiency in early and late bilingualism," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 15–35. doi: 10.1017/CBO9781107375345.002
- Nicoladis, E., Palmer, A., and Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Dev. Sci.* 10, 237–254. doi: 10.1111/j.1467-7687.2007.00582.x
- Nicoladis, E., and Paradis, J. (2011). Learning to liaise and elide comme il faut: evidence from bilingual children. *J. Child Lang.* 38, 701–730. doi: 10.1017/S0305000910000231
- Paradis, J. (2011). Individual differences in child English second language acquisition. *Linguist. Approaches Biling.* 1, 213–237. doi: 10.1075/lab.1.3.01par
- Paradis, J., Nicoladis, E., Crago, M., and Genesee, F. (2011). Bilingual children's acquisition of the past tense: a usage-based approach. *J. Child Lang.* 38, 554–578. doi: 10.1017/S0305000910000218
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Appl. Psycholinguist.* 28, 399–410. doi: 10.1017/S014271640707021X
- 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
- Place, S., and Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Dev.* 82, 1834–1849. doi: 10.1111/j.1467-8624.2011.01660.x
- 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
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ribot, K. M., Hoff, E., and Burridge, A. (2017). Language use contributes to expressive language growth: evidence from bilingual children. *Child Dev.* 89, 929–940. doi: 10.1111/cdev.12770
- Schmeißer, A., Hager, M., Gil, L. A., Jansen, V., Geveler, J., Eichler, N., et al. (2016). "Related but different: the two concepts of language dominance and language proficiency," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 36–65.
- Serratrice, L., Sorace, A., Filiaci, F., and Baldo, M. (2009). Bilingual children's sensitivity to specificity and genericity: evidence from metalinguistic awareness. *Bilingualism* 12, 239–257. doi: 10.1017/S13667289090004027
- Sheng, L., Lu, Y., and Gollan, T. H. (2014). Assessing language dominance in Mandarin-English bilinguals: Convergence and divergence between subjective and objective measures. *Bilingualism* 17, 364–383. doi: 10.1017/S1366728913000424
- Silva-Corvalán, C., and Treffers-Daller, J. (2016). *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107375345
- Sundara, M., Polka, L., and Molnar, M. (2008). Development of coronal stop perception: bilingual infants keep pace with their monolingual peers. *Cognition* 108, 232–242. doi: 10.1016/j.cognition.2007.12.013
- Thordardottir, E. T. (2011). The relationship between bilingual exposure and vocabulary development. *Int. J. Biling.* 15, 426–445. doi: 10.1177/1367006911403202
- Thordardottir, E. T. (2015). The relationship between bilingual exposure and morphosyntactic development. *Int. J. Speech Lang. Pathol.* 17, 97–114. doi: 10.3109/17549507.2014.923509
- Treffers-Daller, J., and Korybski, T. (2016). "Using lexical diversity measures to operationalise language dominance in bilinguals," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 106–133. doi: 10.1017/CBO9781107375345.006
- Tuller, L., Hamann, C., Chilla, S., Ferré, S., Morin, E., Prevost, P., et al. (2018). Identifying language impairment in bilingual children in France and in Germany. *Int. J. Lang. Commun. Disord.* 53, 888–904. doi: 10.1111/1460-6984.12397
- Unsworth, S. (2005). *Child L2, Adult L2, Child L1: Differences and Similarities. A Study on the Acquisition of Direct Object Scrambling in Dutch*. Ph.D. thesis, Utrecht University, Utrecht.
- 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. (2016a). "Amount of exposure as a proxy for dominance in bilingual language acquisition," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*, eds C. Silva-Corvalán and J. Treffers-Daller (Cambridge: Cambridge University Press), 156–173. doi: 10.1017/CBO9781107375345.008
- Unsworth, S. (2016b). "Quantity and quality of language input in bilingual language development," in *Bilingualism Across the Lifespan*, eds E. Nicoladis and S. Montanari (Washington, DC: American Psychological Association), 103–122.
- Ward, J. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845
- Yip, V., and Matthews, S. (2006). Assessing language dominance in bilingual acquisition: a case for mean length utterance differentials. *Lang. Assess. Q.* 3, 97–116. doi: 10.1207/s15434311laq0302\_2

**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 IF and handling Editor declared their shared affiliation.

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

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [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