THE NOTION OF THE NATIVE SPEAKER PUT TO THE TEST: RECENT RESEARCH ADVANCES

EDITED BY: Pedro Guijarro-Fuentes, Cristina Suarez-Gomez,
Mila Vulchanova, Antonella Sorace and Valentin Vulchanov

PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88974-965-2
DOI 10.3389/978-2-88974-965-2

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

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

THE NOTION OF THE NATIVE SPEAKER PUT TO THE TEST: RECENT RESEARCH ADVANCES

Topic Editors:
Pedro Guijarro-Fuentes, University of the Balearic Islands, Spain
Cristina Suarez-Gomez, University of the Balearic Islands, Spain
Mila Vulchanova, Norwegian University of Science and Technology, Norway
Antonella Sorace, University of Edinburgh, United Kingdom
Valentin Vulchanov, Norwegian University of Science and Technology, Norway

Citation: Guijarro-Fuentes, P., Suarez-Gomez, C., Vulchanova, M., Sorace, A., Vulchanov, V., eds. (2022). The Notion of the Native Speaker Put to the Test: Recent Research Advances. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-965-2

Table of Contents

05 ***Editorial: The Notion of the Native Speaker Put to the Test: Recent Research Advances***
Mila Vulchanova, Valentin Vulchanov, Antonella Sorace, Cristina Suarez-Gomez and Pedro Guijarro-Fuentes

11 ***Moving Beyond the Native-Speaker Bias in the Analysis of Variable Gender Marking***
Aarnes Gudmestad, Amanda Edmonds and Thomas Metzger

24 ***The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry***
Anna-Lena Scherger, Gianna Urbanczik, Timon Ludwigs and Jasmin M. Kizilirmak

43 ***Corrigendum: The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry***
Anna-Lena Scherger, Gianna Urbanczik, Timon Ludwigs and Jasmin M. Kizilirmak

45 ***The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures***
Lauretta S. P. Cheng, Danielle Burgess, Natasha Vernooij, Cecilia Solís-Barroso, Ashley McDermott and Savithry Namboodiripad

67 ***Deconstructing the Native Speaker: Further Evidence From Heritage Speakers for Why This Horse Should Be Dead!***
Wintai Tsehaye, Tatiana Pashkova, Rosemarie Tracy and Shanley E. M. Allen

81 ***Listeners' Linguistic Experience Affects the Degree of Perceived Nativeness of First Language Pronunciation***
Lisa Kornder and Ineke Mennen

97 ***Assignment of Grammatical Gender in Heritage Greek***
Demetris Karayiannis, Maria Kambanaros, Kleanthes K. Grohmann and Artemis Alexiadou

107 ***English Adjectives and Estonian Nouns: Looking for Agreement?***
Daria Bahtina, Helin Kask and Anna Verschik

121 ***The Native Speaker in Italian-Dialects Bilingualism: Insights From the Acquisition of Vicentino by Preschool Children***
Emanuela Sanfelici and Maja Roch

137 ***The 'Comparative Logic' and Why We Need to Explain Interlanguage Grammars***
Laura Domínguez and María J. Arche

151 ***Going Native? Yes, If Allowed by Cross-Linguistic Similarity***
Gillen Martínez de la Hidalga, Adam Zawiszewski and Itziar Laka

165 ***A Challenge for Contrastive L1/L2 Corpus Studies: Large Inter- and Intra-Individual Variation Across Morphological, but Not Global Syntactic Categories in Task-Based Corpus Data of a Homogeneous L1 German Group***
Anna Shadrova, Pia Linscheid, Julia Lukassek, Anke Lüdeling and Sarah Schneider

194 ***Individual Sensitivity to Change in the Lingua Franca Use of English***
Irene Taipale and Mikko Laitinen

209 ***Heritage Speakers as Part of the Native Language Continuum***
Heike Wiese, Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Kateryna Iefremenko, Maria Martynova, Tatiana Pashkova, Vicky Rizou, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Wintai Tsehaye, Sabine Zerbian and Yulia Zuban

228 ***Changes in Native Sentence Processing Related to Bilingualism: A Systematic Review and Meta-Analysis***
Patricia Román and Irene Gómez-Gómez

243 ***On the Reliability of the Notion of Native Signer and Its Risks***
Giorgia Zorzi, Beatrice Giustolisi, Valentina Aristodemo, Carlo Cecchetto, Charlotte Hauser, Josep Quer, Jordina Sánchez Amat and Caterina Donati

frontiers in Psychology

EDITORIAL
published: 29 March 2022
doi: 10.3389/fpsyg.2022.875740

Editorial: The Notion of the Native Speaker Put to the Test: Recent Research Advances

Mila Vulchanova[1]*, *Valentin Vulchanov*[1], *Antonella Sorace*[2], *Cristina Suarez-Gomez*[3] and *Pedro Guijarro-Fuentes*[3]

[1] Norwegian University of Science and Technology, Trondheim, Norway, [2] University of Edinburgh, Edinburgh, United Kingdom, [3] Department of Spanish, Modern and Classic Philology, University of the Balearic Islands, Palma, Spain

Keywords: native speaker, language competence, second language acquisition, psycholinguistics, sociolinguistics and bilingualism

Editorial on the Research Topic

The Notion of the Native Speaker Put to the Test: Recent Research Advances

INTRODUCTION

The notion of the native speaker has occupied a prominent place in foreign/second language research and theoretical linguistics: it has influenced both the way we theorize about language and the way we conduct empirical research, and has had practical implications concerning second language pedagogy. In the second language tradition, it was needed as a norm and a standard to evaluate L2 attainment, and, as such, has functioned as a benchmark in terms of goals for L2 instruction (Davies, 2003). In line with this tradition, second language and lingua franca speakers' achievements have ideally been compared with those of monolingual native speakers, although these constitute different groups of speakers, with different needs and abilities. In a similar vein, much psycholinguistic and bilingualism research uses the "native speaker" norm, based on an idealized first language (L1) competence, adopting the inclusion of a control group of monolingual speakers of the language as default in empirical research.

The idealization of the native speaker has its roots in theoretical linguistics, in all likelihood arising from Bloomfield (1927) stringent criteria, whereby the native speaker is an idealized bearer of L1 competence, thus downplaying individual variation. On Chomsky's early definition (Chomsky, 1965), native speakers are characterized by the ability to provide valid judgments on their language and identify ill-formed grammatical expressions in that language, although they may not be able to explain exactly why they are ill-formed. Other attempts at characterizing the native speaker resort to specific key abilities, such as saying the same thing in different ways, hesitating and using fillers, predicting what the other person is going to say, and adding new verbal skills from mere language experience, such as immersion (Halliday, 1978).

Current understanding has moved away from this idealized characterization of the native/monolingual speaker. One of the main reasons for this is the recognition of the amply documented individual variation in language competence, both in adult native speakers, and across language development (Bates et al., 1995; Dabrowska, 2014; Tomblin and Nippold, 2014). Furthermore, neuroscience research has documented the absence of structural differences in key brain structures underlying language use in monolinguals and bilinguals with a language acquisition onset before 3 years of age (Klein et al., 2013). In addition, the role of age of onset as key factor in language competence, has been confirmed in a recent study (Bylund et al., 2020). We also have evidence that L2 speakers may attain high levels of proficiency even with a late onset

OPEN ACCESS

Edited and reviewed by:
Manuel Carreiras, Basque Center on Cognition, Brain and Language, Spain

***Correspondence:**
Mila Vulchanova
mila.vulchanova@ntnu.no

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

Received: *14 February 2022*
Accepted: *08 March 2022*
Published: *29 March 2022*

Citation:
Vulchanova M, Vulchanov V, Sorace A, Suarez-Gomez C and Guijarro-Fuentes P (2022) Editorial: The Notion of the Native Speaker Put to the Test: Recent Research Advances. Front. Psychol. 13:875740. doi: 10.3389/fpsyg.2022.875740

(Sorace and Filiaci, 2006 on "near-native speakers"; Abrahamsson and Hyltenstam, 2009), and can display sensitivity to properties of the L2, including prediction of upcoming words, in response to mere exposure, in a way similar to monolingual L1 speakers (Treffers-Daller and Calude, 2015). It has also been shown that variation in certain cognitive abilities and competencies can account for exceptional skills at second language learning (Vulchanova et al., 2012a,b; Hyltenstam et al., 2018). Cook (1996), for instance, proposes to replace the notion of ideal speaker with "multi-competent language user." Global varieties of the same language offer systematic differences at all levels of language structure, and further challenge unitary perspectives on a single native speaker standard. Emerging new fields of research, such as heritage language and language attrition, are further challenging earlier perceptions of how native speakers should be defined and are re-defining previous assumptions in the field.

THE EVIDENCE

One of the reasons for revising the notion of the native speaker has been the *multilingual turn* (May, 2014), which has given rise to heightened awareness of the cognitive and neuro-biological consequences of being exposed to more than one language over the lifetime (Kroll and Bialystok, 2013). Another key factor has been mounting evidence and empirical findings suggesting large inter-individual variation in native speakers of specific languages as well as the recognition of different sub-groups or populations which all can be characterized as "native speakers."

Individual Variation

Individual variation in language competence has been amply documented, both in adult native speakers, and across language development (Bates et al., 1995; Mulder and Hulstijn, 2011; Dabrowska, 2012; Tomblin and Nippold, 2014). Thus, Mulder and Hulstijn (2011) provide evidence of variation in both lexical and production skills in a sample of adult native speakers of Dutch as a function of age and level of education and profession. In a similar vein, Street and Dabrowska (2010) provide evidence of variability in quantifier knowledge and competence in the comprehension of passives among young adult native speakers of English contingent on education level. Dabrowska (2012) provides a comprehensive review and discussion of findings in the field of native speaker knowledge in a variety of domains, including inflectional morphology, passives, quantifiers, and syntax. A possible account may be sought in how language learners attend to, and interact with, the input and eventually end up with different grammars, while other differences may be attributed to more varied linguistic experience as a result of education. On this backdrop, Dabrowska (2012) suggests that such findings have consequences for research on bilingualism, ultimate attainment in second language acquisition, as well as important methodological implications for the science of language. Inter- and intra-individual variability in adult native speakers of German is documented in the contributions to the current Research Topic. Shadrova et al. show that the variation in certain linguistic aspects in corpus native speaker data undermines general statements about quantitative expectations in L1. The authors also find differences between the phenomena under study. Thus, while morphological and syntactic sub-classes of verbs and nouns show great variability in their distribution in native speaker writing, other, coarser categories, like parts of speech, or types of syntactic dependencies, behave more predictably and homogeneously.

Children acquiring language are another group of native speakers where large variation can be observed applying to expected language competence at specific ages. Research has documented that milestones in language development are less related to age than to earlier acquired language skills (as well as non-linguistic skills) which tend to scaffold them. This may explain the large variation in early vocabulary size and the dependence of early grammar skills on lexical skills in early development (Bates et al., 1995; Bates and Goodman, 1997). Fenson et al. (1994) provide evidence of extensive variability in the rate of lexical, gestural, and grammatical development in a large sample of infants between 8 and 30 months of age. This variability in both the onset and the course of development challenges the notion of a "model" child. At the same time, however, reliable intercorrelations can be found both concurrently and predictively between specific communication skills indicative of an "internal" causality mechanism, with development contingent on an underlying trajectory.

The evidence from typical language development and its "internal" logic is further confirmed by research on neuro-developmental deficits, such as autism and language impairment. Thus, deictic gestures have been documented as an important and reliable predictor of language status for infants between 15 and 20 months (Colonnesi et al., 2010). However, no relationship can be established later, as other communication skills which are more directly related to language take over as predictors. In contrast, for infants and children with autism, pointing gestures continue to exert a strong predictive relationship also at later stages, as shown by Ramos-Cabo et al. (2022). Neuro-developmental deficits also highlight the fact that not all children, who, on other criteria may be considered native speakers, achieve uniform ultimate attainment in the first language, either as a result of impairment in the mechanisms which underlie language acquisition, such as the phonological loop of the memory system or impairment in the mechanisms which ensure efficient language use, e.g., phonological processing problems or attention deficits (Bishop, 2009).

Brain and Age

Neuroscience research provides further evidence challenging the assumed differences between monolingual native speakers and bilinguals exposed to more than one language during the first years of life. Thus, Klein et al. (2013), document that there are no structural differences in key brain structures underlying language use in monolinguals and bilinguals with a language acquisition onset before 3 years of age. The role of age of onset as key factor in language outcomes has been systematically confirmed (Bylund et al., 2020). Recent mounting evidence from bilingual children suggests incomplete acquisition or even loss (Ventureyra, 2004) when exposure to input is terminated before ultimate attainment. There is also evidence of an advantage

in processing or learning later in life the language which individuals were originally exposed to in early infancy, e.g., in the case of internationally adopted children for whom exposure to the birth language has been discontinued (Hyltenstam et al., 2009; Oh et al., 2010). Interestingly, the birth language of internationally adopted children leaves a trace in the neural organization of their language system, and research documents influence of early experience on later brain outcomes. In a study of internationally adopted children (aged 9–17 years), who were completely separated from their birth language (Chinese) at 12.8 months of age, Pierce et al. (2014) show, that, on average, these children displayed brain activation to Chinese linguistic elements that precisely matched that of native Chinese speakers, despite the fact that the international adoptees had no subsequent exposure to Chinese and no conscious recollection of that language. Crucially, no similar activation was found in a control (monolingual) group of speakers of French, the first language of the adopted children. Such findings further raise questions about the nature of, and criteria on which we identify, native speakers. This study and other studies of children with delayed or qualitatively compromised input (e.g., children born profoundly deaf and exposed to oral language following cochlear implantation and internationally adopted children who have delayed exposure to the adoption language; children who experience impoverished language input, that is, children who experience early bouts of otitis media and signing deaf children born to non-signing hearing parents) demonstrate that language outcomes in a first language are the result of an intricate and dynamic interaction of a number of factors, some internal to the child (e.g., phonological working memory) and external factors (e.g., the quantity and quality of the input) (Pierce et al., 2017).

First and Second Language Speakers

Evidence from the SLA research has further questioned the notion of the native speaker. Thus, some advanced second language speakers with a late onset have been shown to perform similarly to participants with native speaker competence (Abrahamsson and Hyltenstam, 2009), although not in all respects: some grammatical properties requiring "interfacing" conditions of a non-linguistic nature remain variable in very advanced levels of L2 competence (Sorace and Filiaci, 2006; Tsimpli and Sorace, 2006; Sorace, 2011). In addition, sensitivity to properties of the target language (an L2 for participants in that research), such as, e.g., prediction of upcoming words and language structure has been documented in L2 learners in response to exposure, in a way similar to monolingual L1 speakers (Treffers-Daller and Calude, 2015). The contribution by Hidalga et al. explores the factors which may facilitate the acquisition of another language to proficiency levels comparable with monolingual native competence. In an EEG experiment, they show that grammar phenomena which are shared by the two languages of early and proficient bilinguals are more likely to be acquired to native level and elicit similar brain responses in both groups of speakers. Furthermore, variation in certain cognitive abilities and competences have been found to account for exceptional skills at second language learning (Hyltenstam et al., 2018). Vulchanova et al. (2012a) document German language skills similar to age-matched children acquiring German as a native language in a child with exceptional skills at learning foreign languages, who learned German from mere exposure to a German TV channel.

The prevalence of multilingualism on a global level has also given impetus to re-conceptualisations of the notion of the native speaker. Cook (1996, 1999) proposes to replace the notion of "ideal speaker" by the notion of "multi-competent language user." In addition, global varieties of the same language (such as e.g., English) offer systematic differences at all levels of language structure, and further challenge unitary perspectives on a single native speaker standard. A similar concept, from the field of English as a Lingua Franca (ELF), is that of "similect," coined by (Mauranen, 2012: 29), to refer to a linguistic variety spoken by people with different L1s, with features transferred from the L1 by individual speakers.

In the field of World Englishes, second language varieties which emerged after a process of language contact between English and the relevant local languages (e.g., Indian English, Jamaican Englishes, etc.) are also relevant in this respect, where the concept of native speaker also needs to be recontextualized. Here, the traditional concept of the native speaker as an idealized monolingual speaker has to be adapted to these multilingual contexts and speakers who speak native varieties which differ from what has traditionally been considered the norm (Mesthrie, 2010). These native varieties have been shaped after a second language acquisition process in language contact contexts and in a globalized world, and determined by linguistic forces such as the increasing use of isomorphic structures which aim at communicative efficiency.

Language Attrition, Heritage Languages, Signers

Defining the native speaker is even more problematic in the face of language attrition and the increase in number of heritage language speakers. A distinction has to be made between attrition in first-generation L1 speakers and inter-generational attrition in heritage speakers, and the connection between the two is currently being explored (see, e.g., Sorace, 2016). A change in native language competence as a result of decreased exposure to the first language is increasingly becoming common in a global world with increased population mobility. Language attrition is manifested at all levels and has been documented particularly at the lexical and syntactic levels (see Schmid and Köpke, 2019), but can also be manifested at the level of processing (see Roman and Gómez-Gómez's contribution in this special issue), and phonology. In their contribution, Kornder and Mennen provide evidence that native speakers of Austrian German who have resided over a long period in and English-speaking environment sound less native to naïve Austrian German monolingual judges. Interestingly, however, this judgement differed significantly from that of a group of naïve Austrian German-English bilinguals, who rated the target group as more native-sounding.

The contribution by Wiese et al. investigated heritage Greek, Russian, Turkish, and German in comparison to monolingual, non-heritage speakers and found non-canonical patterns not

only in bilingual, but also in monolingual speakers, including patterns that have so far been considered absent from native grammars, in domains of morphology, syntax, intonation, and pragmatics. This study also confirms other findings of monolingual heterogeneity and reports a degree of lexical and morphosyntactic inter-speaker variability in monolinguals, sometimes higher than that of bilinguals, further challenging the model of the streamlined native speaker. Also, in some respects, the monolingual participants and the heritage speakers performed similarly.

Tsehaye et al. address further problems with applying the notion of the native speaker in the context of heritage speakers. In this contribution to the Research Topic, they provide evidence by focusing both on similarities and differences between heritage speakers and monolingually-raised speakers, respectively, in their heritage and majority languages. Heritage speakers are an interesting case to study since they are bilinguals who acquire a family (heritage) language, often from parents who are undergoing attrition, and a societal (majority) language in early childhood. In such a way, naturalistic exposure from early childhood qualifies them as native speakers of their heritage language. In addition, some heritage speakers are simultaneous bilinguals, which makes them native speakers of their majority language as well. Others are early second language acquirers who may be indistinguishable from simultaneous bilinguals. Thus, heritage language, while being a challenge for traditional assumptions, also provides a suitable test-bed for these notions. It is clear that research on L1 attrition, within and across generations, flags up important issues in our understanding of the notion of "native speaker" not only of bilingualism and language learning, but also—more generally—in linguistics research. Data showing a convergence between L1 speakers' attrition and advanced L2 speakers' acquisition (Sorace, 2011, 2016) pave the way to the hypothesis that L1 changes may be functional to successful L2 learning: this further undermines the use of the monolingual native speaker as the point of reference, both in research and in society.

Further issues with the notion arise in the context of sign language users depending on the circumstances of acquisition. In their contribution to this Research Topic, Cheng et al. argue against applying the notion in sign language research and psycholinguistics, because it has been inconsistently conceptualized. In addition, factors, such as age, order, and context of acquisition, in addition to social/cultural identity, are often differentially conflated. Zorzi et al. argue further that, given that around 95% of deaf infants are born into a hearing family, deaf signers are exposed to a sign language at various moments of their life, and not only from birth. Since the linguistic input these children are exposed to is not always a fully-fledged natural sign language, the notion of native signer as someone exposed to language from birth cannot be applied. In their contribution, the authors present the results of the first large-scale cross-linguistic investigation on the effects of age of exposure to sign language in each of three sign languages (Catalan Sign Language, French Sign Language, and Italian Sign Language). This study shows the importance of exposure from birth as a "nativeness" criterion, in so far sign language is concerned, with significant differences across language and tests between signers exposed to sign language from birth and those exposed in the first years of life, at least for syntax. On the backdrop of their results across the three different groups, the authors further argue against the generalized use of native signer's grammar as the baseline for language description and language assessment.

CONTRASTING VIEWS

The evidence reviewed above indicates that questioning the notion of the native speaker has gained recognition from a variety of perspectives and fields, including the difficulties to define and operationalize it, its possible bias toward monolingualism, and its potential application to exclusionary purposes (Dewaele, 2018). The traditional definitions of what "first language" means is today at stake because of the evidence of numerous forms of initial bilingualism (Grosjean, 2010) and by the understanding that speakers themselves modify it (Seals, 2019). The negative consequences of applying native speaker standards have been observed also in second/foreign language pedagogy. For instance, *native-speakerism* (the adherence to an (idealized) monolingual target language standard) has been defined as:

> "(…) a neoracist ideology that has wide-ranging impact on how teachers are perceived by each other and by their students. By labeling teachers as separate "native speakers" and "non-native speakers," it falsely positions them as culturally superior and inferior with separate roles and attributes" (Holliday, 2018: 1).

Waddington (2021) provides evidence that the "ideal native speaker" model prevails in pre-service teacher assumptions and beliefs and that the latter not only serve to perpetuate the ideal itself, but also reinforce disempowering and discriminatory attitudes among the profession, which are both outdated and incongruent with current multilingualism inspired policies in early childhood education.

The notion of the native speaker has thus engendered contrastive, but not incompatible views. While some call for completely removing the notion from research, language theory and language pedagogy (see Dewaele, 2018 for an overview), others are proposing critical and well-argued re-conceptualisations. Among the first articulate proposals in that respect is Escudero and Sharwood Smith (2001). They propose a graded notion based on Rosh's prototype theory whereby some speakers, on specific criteria and under certain circumstances will count as native speakers. The criteria these authors propose are intra-linguistic (language competence proper) and extra-linguistic (initial and later language environment, education and literacy). The extra-linguistic aspect has been highlighted in Cook (1999) and Davies (2003) who suggest that language acquisition in naturalistic circumstances is one of the hallmarks of the concept of the native speaker. Importantly, Escudero and Sharwood Smith (2001) distinguish between applying those criteria by naïve (native) speakers of the language and linguists. Based on his research and applying the Shared/Basic Language Cognition framework, Hulstijn (2019) proposes

two ways of defining native speakers in terms of language cognition: in terms of shared/basic language cognition and non-shared/extended/higher language cognition. He also proposes ways of defining native speakers in extralinguistic terms along (a) the biographical/ecological dimension of degrees of being bilingual and (b) the dimension of literacy. Thus, it is being argued that differences in native speakers' language cognition can be primarily conceptualized as a function of their memberships along the extralinguistic dimensions.

New Speakers in Demo-Linguistics: Language Revitalization and Language Maintenance. Notion Still Useful?

Despite theoretical refinement and empirical evidence challenging the native speaker concept, it continues to be used, especially in socio- and demo-linguistics. Although its application may vary, language censuses and ethnolinguistic surveys often include it to refer to groups of people who acquired the same language(s) with their family of origin (Humbert et al., 2018). This perseverance could be probably explained because, in quantitative terms, the native speaker concept may still be applicable. Study after study corroborate that, in multilingual societies, there exists a strong correlation between the condition of being a first language speaker of a given language and scoring higher than speakers of other languages in terms of language proficiency, as well as from the point of view of language use, language dominance, and identification with the language. Needless to say, this strong association does not mean that native-speakerness is always the best predictor of linguistic performance, competence, or attitudes, nor does it allow for ecological fallacy, because the characteristics of individuals are most of the time not determined by the group they belong to.

Further evidence in defense of the notion comes from the fields of language revitalization (Hornsby, 2015; O'Rourke and Ramallo, 2015; Glinert, 2017), language survival and language maintenance, where a vibrant community of native speakers can safeguard and maintain a threatened language (Fishman, 1991; UNESCO, 2003; Phaidin and Cearnaigh, 2008; Giollagain, 2014a,b).

CONCLUDING REMARKS

This Research Topic has aimed to engender a discussion of the notion of the native speaker, from a theoretical point of view, and informed by empirical findings.

Individual contributions introduce important methodological advances in the field by elaborating design and sample standards for future research in bilingualism, heritage language and language attrition (Duñabeitia and Carreiras, 2015; Paap et al., 2015). The current contributions are from the fields of psycholinguistics, sociolinguistics, second language learning, bilingualism, heritage languages, language attrition, and present diverse language learning and acquisition scenarios and languages/linguistic phenomena from different fields, thus encouraging interdisciplinary research. The evidence is rich, but not controversial. It is also unified in suggesting that a reconceptualization of the notion of the native speaker is mandatory for the purposes of language theory, empirical research and language second language instruction.

AUTHOR CONTRIBUTIONS

MV, VV, AS, CS-G, and PG-F wrote the manuscript and approved the final version. All authors contributed to the article and approved the submitted version.

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

Bates, E., Dale, P., and Thal, D. (1995). "Individual differences and their implications for theories of language development," in *The Handbook of Child Language*, eds P. Fletcher, and P. B. MacWhinney (London: Blackwell). 95–151. doi: 10.1111/b.9780631203124.1996.00005.x

Bates, E., and Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: evidence from acquisition, aphasia and real-time processing. *Lang. Cogn. Process.* 12, 507–584. doi: 10.1080/016909697386628

Bishop, D. (2009). Specific language impairment as a language learning disability. *Child Lang. Teach. Ther.* 25, 163–165. doi: 10.1177/0265659009105889

Bloomfield, L. (1927). "Literate and Illiterate speech," in *Language in Culture and Society*, ed D. Hymes (New York, NY: Harper and Row), 391–396.

Bylund, E., Hyltenstam, K., and Abrahamsson, N. (2020). Age of acquisition - not bilingualism - is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism* 24, 1–13. doi: 10.1017/S1366728920000188

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. doi: 10.21236/AD0616323

Colonnesi, C., Stams, G., Koster, I., and Noom, M. J. (2010). The relation between pointing and language development. *Dev. Rev.* 30, 352–366. doi: 10.1016/j.dr.2010.10.001

Cook, V. (1996). "Competence and multi-competence," in *Performance and Competence in Second Language Acquisition*, eds G. Brown, K. Malmkjaer, and J. Williams (Cambridge: CUP), 57–69.

Cook, V. (1999). Going beyond the Native Speaker in Language Teaching. *TESOL Quart.* 33, 185–209. doi: 10.2307/3587717

Dabrowska, E. (2012). Different speakers, different grammars: individual differences in native language attainment. *Linguistic Approaches Bilingual.* 2, 219-253. doi: 10.1075/lab.2.3.01dab

Dabrowska, E. (2014). 'Words that go together': measuring individual differences in native speakers' knowledge of collocations. *Mental Lexicon* 9: 401–418. doi: 10.1075/ml.9.3.02dab

Davies, A. (2003). *The Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters. doi: 10.21832/9781853596247

Dewaele, J.-M. (2018). Why the dichotomy 'L1 Versus LX User' is better than 'native versus non-native speaker'. *Appl. Linguist.* 32, 236–240. doi: 10.1093/applin/amw055

Duñabeitia, J. A., and Carreiras, M. (2015). The bilingual advantage: acta est fabula? *Cortex* 73, 371–372. doi: 10.1016/j.cortex.2015.06.009

Escudero, P., and Sharwood Smith, M. (2001). Reinventing the native speaker: Or "What you never wanted to know about the native speaker so never dared to ask." *EUROSLA Yearbook* 1, 275-286. doi: 10.1075/eurosla.1.20esc

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., and Pethick, S. (1994). Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* 59, 1–173. doi: 10.2307/1166093

Fishman, J. A. (1991). *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Bristol: Multilingual Matters.

Giollagain, C. O. (2014a). Unfirm ground: a re-assessment of language policy in ireland since independence. *Lang. Prob. Lang. Plann.* 38, 19–41. doi: 10.1075/lplp.38.1.02gio

Giollagain, C. O. (2014b). From revivalist to undertaker: new developments in official policies and attitudes to Ireland's 'First language. *Lang. Prob. Lang. Plann.* 38, 101–127. doi: 10.1075/lplp.38.2.01gio

Glinert, L. (2017). *The Story of Hebrew*. Princeton, NJ: Princeton University Press.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459

Halliday, M. A. K. (1978). *Language as Social Semiotic*. London: Edward Arnold.

Holliday, A. (2018). ≪Native-Speakerism,≫ in *The TESOL Encyclopedia of English Language Teaching*, ed J. I. Liontas (New York, NY: Wiley). doi: 10.1002/9781118784235.eelt0027

Hornsby, M. (2015). *Revitalizing Minority Languages: New Speakers of Breton, Yiddish and Lemko*. London: Palgrave Macmillan UK. doi: 10.1057/9781137498809

Hulstijn, J. H. (2019), An individual-differences framework for comparing nonnative with native speakers: perspectives from BLC theory. *Lang. Learn.* 69, 157–183. doi: 10.1111/lang.12317

Humbert, P., Coray, R., and Duchene, A. (2018). *Compter Les Langues: Histoire, Méthodes et Politiques Des Recensements de Population*. Une Revue de La Littérature. Fribourg/Freiburg: Institut de plurilinguisme.

Hyltenstam, K., Bartning, I., and Fant, L. (eds.). (2018). "Polyglotism: a synergy of abilities and predispositions," in *High-Level Language Proficiency in Second Language and Multilingual Contexts* 170–195. doi: 10.1017/9781316809686.008

Hyltenstam, K., Bylund, E., Abrahamsson, N., and Park, H.-S. (2009). Dominant-language replacement: the case of international adoptees. *Bilingualism* 12, 121–140. doi: 10.1017/S1366728908004008

Klein, D., Mok, K., Chen, J. K., and Watkins, K. E. (2013). Age of language learning shapes brain structure: a cortical thickness study of bilingual and monolingual individuals. *Brain Lang.* 131, 20–24. doi: 10.1016/j.bandl.2013.05.014

Kroll, J., and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *J. Cogn. Psychol.* 25, 10.1080/20445911.2013.799170. doi: 10.1080/20445911.2013.799170

Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-Native Speakers*. Cambridge: Cambridge University Press.

May, S. (Ed.) (2014). *The Multilingual Turn: Implications for SLA, TESOL, and Bilingual Education*. New York, NY: Routledge.

Mesthrie, R. (2010). New englishes and the native speaker debate. *Lang. Sci.* 32, 594–601. doi: 10.1016/j.langsci.2010.08.002

Mulder, K., and Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Appl. Linguist.* 32, 475–494. doi: 10.1093/applin/amr,016

Oh, J. S., Au, T. K.-F., and Jun, S.-A. (2010). Early childhood language memory in the speech perception of international adoptees. *J. Child Lang.* 37,1123–1132 doi: 10.1017/S0305000909990286

O'Rourke, B., and Ramallo, F. (2015). Neofalantes as an active minority: understanding language practices and motivations for change amongst new speakers of Galician. *Int. J. Soc. Lang.* 231, 147–165. doi: 10.1515/ijsl-2014-0036

Paap, K., Myuz, H., 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

Phaidin, C. N., and Cearnaigh, S. O. (2008). *A New View of the Irish Language*. Dublin: Cois Life Teoranta.

Pierce, L., Genesee, F., Delcenserie, A., and Morgan, G. (2017). Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Appl. Psycholinguist.* 38, 1265–1300. doi: 10.1017/S0142716417000236

Pierce, L. J., Klein, D., Chen, J. K., Delcenserie, A., and Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proc. Nat. Acad. Sci. U.S.A.* 111, 17314–17319. doi: 10.1073/pnas.1409411111

Ramos-Cabo, S., Acha, J., Vulchanov, V., and Vulchanova, M. (2022). You may point, but do not touch: impact of gesture-types and cognition on language in typical and atypical development. *Int. J. Lang. Commun. Disord.* 57, 324–339. doi: 10.1111/1460-6984.12697

Schmid, M., and Köpke, B. (eds.). (2019). *The Oxford Hanbook of Language Attrition*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198793595.001.0001

Seals, C. A. (2019). *Choosing a Mother Tongue: The Politics of Language and Identity in Ukraine*. Bristol: Multilingual Matters. doi: 10.21832/9781788925006

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguist. Approaches Bilingual.* 1, 1–33. doi: 10.1075/lab.1.1.01sor

Sorace, A. (2016). Referring expressions and executive functions in bilingualism. *Linguist. Approaches Bilingual.* 6, 669–684. doi: 10.1075/lab.15055.sor

Sorace, A., and Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Res. 22* 3, 339–368. doi: 10.1191/0267658306sr271oa

Street, J., and Dabrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120, 2080–2094. doi: 10.1016/j.lingua.2010.01.004

Tomblin, B., and Nippold, M. (2014). *Understanding Individual Differences in Language Development across the School Years. Language and Speech Disorders.* New York, NY: Psychology Press, Taylor and Francis Group. doi: 10.4324/9781315796987

Treffers-Daller, J., and Calude, A. (2015). The role of statistical learning in the acquisition of motion event construal in a second language. *Int. J. Biling. Educ. Biling.* 18, 602–623. doi: 10.1080/13670050.2015.1027146

Tsimpli, I.-M., and Sorace, A. (2006). "Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena," in *Proceedings of the 30th annual Boston University Conference on Language Development (BUCLD)*, eds. D. Bamman, T. Magnitskaia and C. Zaller (Somerville, MA: Cascadilla Press), 653–664.

UNESCO (2003). *Language Vitality and Endangerment*. Available online at: http://www.unesco.org/new~/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language_vitality_and_endangerment_EN.pdf (accessed November 22, 2021).

Ventureyra, V. (2004). The loss of first language phonetic perception in adopted Koreans. *J. Neurolinguistics* 17, 79–91. doi: 10.1016/S0911-6044(03)00053-8

Vulchanova, M., Talcott, J. B., Vulchanov, V., and Stankova, M. (2012b). Language against the Odds, or rather not: the weak central coherence hypothesis and language. *J. Neurolinguistics* 25, 13–30. doi: 10.1016/j.jneuroling.2011.07.004

Vulchanova, M. D., Talcott, J., Brent, Vulchanov, V., Stankova, M., and Eshuis, R. (2012a). Morphology in autism spectrum disorders: local processing bias and language. *Cogn. Neuropsychol.* 29, 584–600. doi: 10.1080/02643294.2012.762350

Waddington, J. (2021). Rethinking the 'ideal native speaker' teacher in early childhood education. *Lang. Cult. Curricul.* 35, 1–17. doi: 10.1080/07908318.2021.1898630

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Communication

ORIGINAL RESEARCH
published: 09 August 2021
doi: 10.3389/fcomm.2021.723496

Moving Beyond the Native-Speaker Bias in the Analysis of Variable Gender Marking

Aarnes Gudmestad[1]*, *Amanda Edmonds*[2] *and Thomas Metzger*[3]

[1]*Virginia Polytechnic Institute and State University, Department of Modern and Classical Languages and Literatures, Blacksburg, VA, United States,* [2]*Université Paul-Valéry Montpellier 3, Département D'études Anglophones, Montpellier, France,* [3]*The Ohio State University, Department of Statistics, Columbus, OH, United States*

In the current study, we respond to calls for reform in second language acquisition that center on the field's preoccupation with native-speaker and prescriptive targets as a benchmark for additional-language learning. In order to address these concerns, we examine the use and development of grammatical gender marking in additional-language Spanish in a prescriptive-independent manner. Specifically, we depart from previous analyses that have centered on accuracy and targetlikeness and we shift the object of analysis to the linguistic forms (i.e., feminine and masculine modifiers) that additional-language participants use. We adopt a variationist approach to explain how participants vary their use of modifier gender and how this use changes longitudinally. We argue that such an approach to studying additional languages allows us to offer new insights about the acquisition of grammatical gender marking in additional-language Spanish. We end by critically reflecting on some of the challenges that we encountered in trying to integrate this paradigm shift into the examination of a well-studied grammatical structure.

Keywords: native speaker bias, variationist approaches, grammatical gender, Spanish, variation

OPEN ACCESS

Edited by:
Pedro Guijarro-Fuentes, University of the Balearic Islands, Spain

Reviewed by:
Jonathan R. Kasstan, University of Westminster, United Kingdom
Ana de Prada Pérez, Maynooth University, Ireland

***Correspondence:**
Aarnes Gudmestad
agudmest@vt.edu

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

Received: *10 June 2021*
Accepted: *26 July 2021*
Published: *09 August 2021*

Citation:
Gudmestad A, Edmonds A and Metzger T (2021) Moving Beyond the Native-Speaker Bias in the Analysis of Variable Gender Marking. Front. Commun. 6:723496. doi: 10.3389/fcomm.2021.723496

INTRODUCTION

Scholars in applied linguistics have raised concerns about the native-speaker bias that has shaped the field. Ortega (2014), p. 32 writes:

> The bias results from the assumption that monolingualism is the default for human communication and from valuing nativeness as a superior form of language competence and the most legitimate relationship between a language and its users. These critiques are poignant in unmasking deeply negative consequences for research and praxis . . . In many of these critiques, the field of second language acquisition (SLA) has been targeted explicitly as suffering in its very core, and in particularly acute ways, from the ailments that result from taking nativeness and monolingualism as natural organizing principles for the study of additional-language learning.

Following these criticisms, Cook and Wei (2016), The Douglas Fir Group (2016), Ortega (2013), Ortega (2017), among others, have advocated for conceptual and methodological reform regarding the role of native-speaker targets in SLA. In the current study, we respond to the call for reform by offering a concrete example of how additional-language[1] data may be profitably analyzed in a prescriptive-independent manner. We conduct an analysis of the development of grammatical

[1]We use this term to refer to any language learned after the first language (L1, cf. The Douglas Fir Group, 2016).

gender marking behavior in additional-language Spanish that crucially does not use a native-speaker or prescriptive norm[2] as a benchmark against which learner behavior is measured. Namely, rather than analyze accuracy in gender marking, we focus on the forms that participants use – namely, feminine and masculine modifiers – and we use a variationist approach to model the variability in the use of these forms at three points over a 21-month period. We conclude not only with a discussion of the specific ways in which SLA stands to benefit from explanations of developmental trajectories that are independent from considerations of a native-speaker target but also with a reflection on the challenges that come with this paradigm shift.

BACKGROUND

The Native-Speaker Norm in Second Language Acquisition

For years now, many researchers working within SLA have been criticizing the field's preoccupation with native-speaker targets (e.g., Bley-Vroman, 1983; Cook, 1992; Cook, 1997; Cook, 2016; The Douglas Fir Group, 2016; Kinginger, 2009; Klein, 1998; Ortega, 2014; Ortega, 2016; Ortega, 2017; see also; Bachman, 1990; Mauranen, 2012).[3] Studies using such targets tend to conduct error analyses or other assessments of accuracy, nativelikeness, or targetlikeness that involve comparing additional-language users to native speakers or prescriptive norms. Some of the arguments against the native-speaker bias are that it implicitly takes a deficit view of additional-language speakers and that it ignores the reality of multilingualism, because "when researchers and educators insist on a monolingual native-speaking golden rule for their interpretations of development, progress, or success, they are setting up L2 [second-language] learners for failure, since multilingual competence is simply different in nature from monolingual competence" (The Douglas Fir Group, p. 35; cf. Hall et al., 2006). In an earlier critique, Bley-Vroman called the comparison of additional-language users to native speakers of the target language the "comparative fallacy" and asserted that it negatively impacts descriptions of additional-language users' linguistic behavior (p. 2). Bachman warned against evaluating additional-language behavior based on native speakers because "native speakers show considerable variation in ability" (p. 39; see also Dabrowska (2012) and Mulder and Hulstijn (2011), for research that offers evidence of variability in linguistic knowledge, oral proficiency, etc. among native speakers). Moreover, another concern about the impact that the native-speaker bias has had on SLA is that because of this bias, SLA has little standing in the language sciences. Klein argued that with a focus on "learners' utterances as deviations from a certain target, instead of genuine manifestations of underlying language capacity … [SLA] analyses them in terms of what they are not rather than what they are" (p. 527). The consequence of this focus is that the observations that emerge from SLA research are not essential to theoretical advancements in linguistics more generally (Klein, 1998, p. 530).

Despite the concerns raised by various leaders in the field with the native-speaker bias in SLA, comparisons of additional-language data to native-speaker or prescriptive baselines remain commonplace (Ortega, 2016). Ortega (2014) has advocated for the need to "replace SLA's existing research goal of explaining why late bi/multilinguals are not native speakers … with the goal of understanding the process and consequences of becoming bilingual or multilingual later in life" (p. 33). This conceptual shift arguably necessitates methodological changes. Perhaps the best-known proposal for navigating this change has come from Cook (e.g., 2016), whose multicompetence approach is offered as a concrete way to carry out SLA research without referencing a native or prescriptive norm. One of the hallmarks of this approach is the need to study the full linguistic repertoire of bi/multilingual speakers – not only any additional languages that speakers may use, but also their L1. Thus, the multicompetence approach necessitates a dramatic shift in research design, which may in part explain why it has not yet been fully embraced in the field. Alongside proposals for more radical design changes, we believe that there is great potential to move beyond the preoccupation with prescriptive and native-speaker comparisons by drawing on frameworks currently in place within SLA. In this vein, we note a call for reform by Ortega (2017) who welcomes "empirical research on all alternatives … because it will make us unearth new knowledge about L2 development" (see also Ortega, 2016). Thus, the overarching goal of the current study is to respond to this call for conceptual and methodological reform with an analysis that takes advantage of one theoretical framework that has been fruitfully applied to SLA research, namely variationism. As we will detail in the following section, the variationist approach offers many strengths that make it compatible with the current endeavor. In the present study, we draw on certain aspects of variationism to study grammatical gender in Spanish, a linguistic phenomenon that to our knowledge has yet to be investigated without comparing additional-language users to a native-speaker target or prescriptive baseline. The analysis we offer is a reanalysis of the data examined in Gudmestad et al. (2019), in which grammatical gender marking was analyzed with reference to a prescriptive norm. With this reanalysis, we hope to show how existing tools in SLA can be valuable in helping the field to fully realize this conceptual and methodological shift.

Variationist Second Language Acquisition

Variationist sociolinguistics is an area of scholarship that focuses on variation and change in language (Labov, 1966). Variable structures, which refer to cases where a single language function

[2]A native-speaker target can either refer to an idealized norm or a target that reflects real-world language behavior of native speakers, whereas a prescriptive norm refers to an idealized norm. In the current article, we use both terms because the native-speaker target and prescriptive norm for grammatical gender marking in Spanish, the linguistic phenomenon under investigation, are largely identical (cf. Gudmestad et al., 2019).

[3]Research on language revitalization and new speakers has voiced similar concerns (e.g., O'Rourke et al., 2015) and addressed issues such as speaker motivation and identity construction (e.g., Nance et al., 2016).

can be expressed by two or more forms, are the object of study. One example is subject expression in Spanish, where a speaker can use either a pronoun (*yo*) or an unexpressed subject to express the first-person singular (see example 1). Researchers typically use multivariate and quantitative analyses in order to identify the multitude of linguistic and social (extra-linguistic) factors that influence the variable occurrence of a given form. Returning to subject expression in Spanish, de Prada Pérez (2015) examined first-person subjects in a native-speaker dataset. When she analyzed all of her data together,[4] she found that three linguistic factors and two social factors simultaneously predicted the use of first-person subjects. For example, the linguistic factor of verb form ambiguity[5] was significant. The probability of using a personal pronoun was higher with ambiguous verbs, whereas the probability of using an unexpressed subject was higher with unambiguous verbs. The results for the social factor of age showed that older speakers favored the use of *yo* and younger speakers favored the use of unexpressed subjects.

1) *Yo no tengo tiempo.* 'I don't have time.'
   ∅ *no tengo tiempo.* '(I) don't have time.'

As an extension of this work, variationist SLA "explores the relationship between contextual variables (both social and linguistic contextual variables) and variation in the form of learner language" (Tarone, 2007, p. 845). Given its connection with sociolinguistics, most variationist scholarship in SLA has examined linguistic phenomena that are sociolinguistically variable among native speakers (e.g., Mougeon et al., 2010). This has been called Type 2 variation (Rehner, 2002). However, this approach has also been adopted to study linguistic structures for which learners exhibit variability but native speakers do not (i.e., Type 1 variation, Rehner, 2002), such as preposition + article contractions in Portuguese (Picoral and Carvalho, 2020) and plural marking in English (Young, 1991). Researchers often examine cross-sectional (e.g., Kanwit, 2017) or longitudinal (e.g., Regan et al., 2009) data to understand how additional-language learners' variable behavior changes along the developmental trajectory. Because there exists a substantial body of variationist SLA research (cf. Geeslin and Long, 2014), we offer an example of a study that used this approach to illustrate what a variationist SLA analysis looks like and to demonstrate what such an analysis of learner data can contribute to SLA.

In Geeslin et al. (2012), the researchers examined the development of perfective past-time reference (i.e., the variation between the preterit and the present perfect to express reference to the perfective past) among additional-language learners who studied in Spain for 7 weeks and a group of native speakers of Peninsular Spanish (a case of Type 2 variation). They analyzed data from a written contextualized task in which participants selected whether they liked the use of the preterit, present perfect, or both forms in specific contexts. The additional-language learners completed the task three times (during weeks 1, 4, and 7 of the study-abroad program) and the native speakers completed it once. Although each item on the written contextualized task presented learners with three response options (preterit, present perfect, both forms), for most of the analysis, the researchers analyzed the present perfect and both responses together as a "present perfect allowed" category. We focus here on the part of the analysis that centered on the predictive factors. In order to examine the factors that predicted the variable selection of perfective verb forms, the researchers performed four regression models – one for each data-collection point for the learners and one for the native speakers. The three learner models enabled the researchers to make observations about additional-language development longitudinally, and assessments of targetlikeness were made by comparing the learner and native-speaker models. One finding, for example, was that years of study, an extra-linguistic variable, impacted verb selection on the task at weeks 1, 4, and 7 for the learners, such that participants who had been studying Spanish for 5 years or more selected more present perfect than those who had studied Spanish for 4 years or fewer. Another result was that the telicity[6] variable, a linguistic factor, was significant for learners at weeks 1 and 4 but not at week 7 or for the native speakers; at weeks 1 and 4 the preterit was more likely than the present perfect in telic contexts.

It is important to note that Geeslin et al. (2012) analyzed native-speaker data, which they then used as a baseline against which to compare learner data. Although the use of native-speaker (or bilingual, e.g., Kanwit, 2017) benchmarks is typical in variationist SLA, we argue that it is not an essential component and that variationism can be fruitfully used to study additional languages without reference to a native or prescriptive norm. Three important characteristics of variationist SLA, all exemplified in Geeslin et al. (2012), show the potential of this approach. The first is that linguistic forms are the object of study; for example, Geeslin et al. (2012) focus on perfective past verb forms. This attention to understanding the occurrence of different forms (instead of, for example, investigating accuracy) makes variationism particularly promising for additional-language research that attempts to avoid comparisons with a prescriptive norm. In this vein, we aim to apply variationist tools in order to analyze the development of grammatical gender marking in additional-language Spanish by focusing on the forms that participants use. We return to the question of the forms that we analyze in the next session.

Second, at the heart of the variationist approach to SLA is the goal of understanding systematic variability in language. In Geeslin et al. (2012), the researchers showed that, rather than selecting a single verb form categorically in certain contexts, the learners' behavior was variable. For example, at weeks 1 and 4, when telicity significantly influenced verb selection, the learners did not select the preterit 100% of the time in telic contexts. Instead, they were more likely to choose the preterit in these contexts but the present perfect was also possible.

[4]She began her analysis by examining four speaker groups together: Spanish L1 bilinguals, Catalan L1 bilinguals, a Spanish control group, and a Catalan control group.

[5]de Prada Pérez (2015), p. 125 coded a verb form "as ambiguous if it was morphologically ambiguous with another [grammatical-person] form".

[6]Telicity refers to whether a predicate has an endpoint (Geeslin et al. (2012) p. 203).

We argue that the focus on variation that characterizes variationist SLA is a strength of this approach that makes it an excellent candidate for attempts to respond to current calls for reform in SLA using established frameworks. The Douglas Fir Group contends, "[v]ariability is not measurement error begging for better control. Acknowledging inter-as well as intra-individual variation helps counter deficit orientations in the description of linguistic development in an L2 . . . and focus on what learners can do rather than what they cannot do" (2016, p. 30). Variationist SLA not only recognizes the presence of variability, it provides conceptual and methodological tools for explaining the complex systematicity and dynamicity of additional-language variation, which leads us to the third characteristic.

The variationist approach to SLA results in detailed observations about the dynamicity of learner language. This is accomplished through the use of multivariate analyses that offer explanations of how the linguistic (e.g., telicity) and extra-linguistic (e.g., years of Spanish study) factors that predict the occurrence of linguistic forms can change over time, thus furthering knowledge about additional-language development. In Geeslin et al. (2012), whereas the impact that years of study had on perfective past reference was stable over time, change in learner behavior was observed with telicity between weeks 4 and 7. Importantly for the current study, both this stability and this change can be observed without relying on a native-speaker baseline. Therefore, in the current study, we aim to examine how the additional-language use of grammatical-gender forms changes (or not) longitudinally. Crucially, however, we do not compare the additional-language data to a native-speaker benchmark or prescriptive norm.

Grammatical Gender in Spanish

In Spanish nouns have either feminine or masculine gender. Gender assignment is arbitrary for most nouns (e.g., *bicicleta*$_{\text{fem}}$ 'bike', *coche*$_{\text{masc}}$ 'car'), though biological sex determines the gender of some nouns (e.g., *hija*$_{\text{fem}}$ 'daughter', *hijo*$_{\text{masc}}$ 'son'). Descriptively, whereas some adjectives and determiners have a single form that is used with nouns of both genders (e.g., *mi* 'my' and *verde* 'green', as in *mi manzana*$_{\text{fem}}$ *verde* 'my green apple' and *mi melón*$_{\text{masc}}$ *verde* 'my green melon'), most have different feminine and masculine forms (e.g., *una*$_{\text{fem}}$ *manzana*$_{\text{fem}}$ *amarilla*$_{\text{fem}}$ 'a yellow apple' and *un*$_{\text{masc}}$ *melón*$_{\text{masc}}$ *amarillo*$_{\text{masc}}$ 'a yellow melon'). Some endings are linked with one gender. The canonical endings for nouns and modifiers are *-a* for feminine and *-o* for masculine. However, there are exceptions, such that nouns ending in *-a* can be masculine (e.g., *poema*$_{\text{masc}}$ 'poem') and those ending in *-o* can be feminine (*nao*$_{\text{fem}}$ 'ship'). There are other endings that are either strongly connected to one gender (e.g., *-tad* as in *libertad*$_{\text{fem}}$ 'freedom' and *-e estante*$_{\text{masc}}$ 'shelf') or that are not linked with a particular gender (e.g., *-s* as in *tos*$_{\text{fem}}$ 'cough' and *mes*$_{\text{masc}}$ 'month'; Teschner and Russell, 1984).

Grammatical gender in additional-language Spanish has been studied extensively, and to our knowledge all of this work has been oriented toward a native-speaker or prescriptive benchmark (cf. Alarcón, 2014). This means that the focus has been on accuracy or targetlikness, where a mismatch in the gender of a noun and its modifier constitutes an error or an instance of non-targetlike behavior. Previous research has sought to better understand how additional-language users of Spanish produce and process gender marking (Alarcón, 2014). We limit our review, however, to production studies since we analyze language use in the present investigation. We focus on the variables identified in this research that explain the development of targetlike grammatical gender marking. Noun gender, noun ending, modifier type, and noun class have been the most widely studied factors in gender-marking research. Findings have shown that learners tend to exhibit more accurate gender marking with masculine nouns (e.g., Finnemann, 1992; White et al., 2004; Montrul et al., 2008), which has been interpreted to indicate that masculine modifiers are a default form that develop more quickly than feminine modifiers. Studies have also shown that nouns that have the prototypical *-o* ending for masculine nouns and *-a* ending for feminine nouns are connected to higher rates of accuracy (e.g., Fernández-García, 1999; Alarcón, 2011). Various investigations (Bruhn de Garavito and White, 2002; White et al., 2004; Alarcón, 2010), though not all (e.g., Montrul et al., 2008; Alarcón, 2011), have found that learners mark gender more accurately on determiners than adjectives (i.e., modifier type). Regarding noun class, studies have demonstrated differing results, with some reporting higher accuracy rates with arbitrary gender (e.g., Bruhn de Garavito and White, 2002) and others showing that learners are more accurate with biological gender (e.g., Fernández-García, 1999). Other factors investigated include noun number, where Finnemann (1992) found that learners exhibited fewer errors with singular compared to plural nouns, and course level or proficiency, about which it was revealed that learners became more accurate with grammatical gender as course level or proficiency increased (e.g., Bruhn de Garavito and White, 2002; Montrul et al., 2008).

Moreover, a recent study examined the development of targetlikeness in grammatical gender marking longitudinally by bringing together the previously studied independent variables and five new factors. In addition to the aforementioned six variables, Gudmestad et al. (2019) analyzed the number of syllables between the noun and the modifier, task (oral interview, oral narration, written essay), time (before study abroad, during study abroad, after study abroad), and two factors that assessed noun frequency: noun log-frequency (language) and noun frequency (individual). The factor noun log-frequency (language) provided a measure of noun frequency in Spanish using the *Corpus del español* (Davies, 2016-), whereas the factor noun frequency (individual) provided a measure of noun frequency that depended on each individual's use (see the *Methods* section for details on the participants, the data collection, and the full set of variables, as the current study constitutes a reanalysis of Gudmestad et al. (2019)). A generalized linear mixed-effects model revealed that noun ending, task, noun gender, noun frequency (individual), syllable distance, modifier type, initial proficiency, time, and the interaction between noun ending and time simultaneously predicted targetlike gender marking in language production. Similar to previous investigations, we found that learners were more likely to be targetlike with gender marking with canonical *-o*/*-a* endings, masculine nouns, and determiners and that higher scores on a proficiency test were also linked to higher rates of targetlike use. Several novel findings came out of this investigation: Learners exhibited higher log-odds of targetlike use on a written essay compared to oral tasks, at the in-stay and post-stay data-collection points compared to the pre-stay time, as the distance between the

noun and the modifier decreased, and as the frequency with which each individual used a noun with a modifier overtly marked for gender increased. Finally, the interaction between noun ending and time indicated that the learners made gains in marking gender with nouns that have what were called "deceptive" endings, that is, nouns that ended in either–*o* or–*a*, but whose gender did not correspond to the canonical gender for that ending (e.g., *poema*$_{masc}$ 'poem', *nao*$_{fem}$ 'ship').

In the current study, we reanalyze our previous work in order to offer a reconceptualization of additional-language data in a prescriptive-independent manner by shifting the focus of analysis from targetlikeness to the forms that participants use. More specifically, instead of analyzing how targetlike learners are in their marking of gender, we analyze their use of feminine and masculine modifiers through the lens of systematic variation. Indeed, gender-marking behavior consists of making a (conscious or unconscious) choice between the masculine and feminine forms of a given modifier; previous research has demonstrated that learners' use of modifier gender is not categorical, which means that certain nouns may be used variably with both feminine and masculine marked modifiers (Dewaele and Véronique, 2001). This variability constitutes a case of Type 1 variation, as gender marking in Spanish (with the exception of certain nouns whose gender varies by geographical region) has not been shown to vary sociolinguistically among native speakers (Gudmestad et al., 2019). Thus, in line with variationist SLA, there is value in modeling variability in the use of these linguistic forms. With the present analysis, we shift the focus from how correct additional-language users are to what predicts their use of modifier gender.

THE CURRENT STUDY

We address the following question in the present investigation: What linguistic and extra-linguistic factors predict the variable use of modifier gender over time? In order to answer this question, we conduct a variationist analysis of gender marking in additional-language Spanish that attempts to be independent from a native-speaker or prescriptive norm. Instead of determining to what extent learners approximate a targetlike norm, the findings contribute insight into factors that predict the use of modifier gender and whether and how these factors change over time. After presenting the results, we discuss the new knowledge about grammatical gender that emerges from this type of analysis, and then we reflect on what an approach to additional-language gender marking that moves away from a native-speaker norm brings to SLA. Specifically, we consider the more general impact and some of the challenges of such an approach on research within the field of SLA.

METHODS

Corpus and Participants

Our data come from LANGSNAP, a publicly available corpus (http://langsnap.soton.ac.uk, e.g., Mitchell et al., 2017). For this corpus the research team collected data from additional-language speakers of French and Spanish over a period of 21 months, which included an academic year abroad in a French- or Spanish-speaking country. The data were collected at six points in time: Before the participants went abroad, three occasions while they were abroad, and twice after returning to the United Kingdom.[7] At each data-collection period, the participants completed an oral interview in which they talked about their lives, an oral picture-based narration task, and a written argumentative essay. We report on half of data-collection periods for the Spanish data. We analyzed all three tasks at three different data-collection periods: Before the participants went abroad (henceforth, pre-stay), the third data collection while abroad that occurred at the end of their stay abroad and 1 year after pre-stay (in-stay), and the final data collection in the United Kingdom that took place 21 months after the initial data collection and about 8 months after returning home (post-stay).

Our dataset consists of 21 of the 27 undergraduate students included in the corpus who were pursuing an undergraduate degree in Spanish in the United Kingdom.[8] They ranged in age from 20 to 25 years ($M = 20.8$, $SD = 1.6$). Fifteen were women and six were men. When the project began, they had been studying Spanish for an average of 5.4 years ($SD = 3.4$, *range* = 2–14). The L1 of the participants was English ($n = 19$), Polish ($n = 1$), or both Polish and English ($n = 1$). In terms of other languages that the participants had studied, 18 indicated that they had studied French, German and/or Italian, two had not learned another language, and one opted not to share this information. During the participants' academic year abroad, they were teaching assistants ($n = 10$), exchange students, ($n = 9$), and workplace interns ($n = 2$), and five lived in Mexico whereas 16 were in Spain.

Data Coding and Analysis

We began the coding by identifying every referent that was modified by a determiner or an adjective ($K = 16{,}357$).[9] The tokens retained for the dataset that we analyzed in the present investigation ($k = 11{,}351$)[10] shared three characteristics: 1) the referent for each token was a noun (instances involving pronouns were coded, but not analyzed in this project), 2) only nouns that occurred more than once in the dataset were analyzed (a total of 482 nouns occurred a single time and were thus removed from our dataset), and 3) each token involves a modifier that exhibited overt gender marking, meaning that the modifier had distinct

[7]The three in-stay periods were collected five, nine and 12 months after the pre-stay data collection. Additionally, see Tracy-Ventura and Huensch (2018) for a presentation of later phases of the project.

[8]We have coded and analyzed a subset of the data-collection points and of the participants because the coding, which was done entirely by hand, was very labor intensive. Additionally, the three data-collection periods that we coded enabled us to make observations about possible change over the course of an academic year abroad (in-stay) and whether any changes held after the participants had returned home (post-stay).

[9]For the oral data, we relied on the transcripts provided online by the LANGSNAP team. After the transcription was completed initially, the transcripts were checked by at least one other member of the LANGSNAP team.

[10]In Gudmestad et al. (2019), our dataset consisted of more observations because we analyzed nouns that occurred once and those that occurred multiple times. This previous analysis did not include a random effect for noun type.

masculine and feminine forms. Examples from the data are available in (2), with the nouns underlined and the modifiers in bold. The examples in (2) illustrate variable use of modifier gender (both feminine and masculine marked modifiers) with a single noun.

2) *Voy a estudiar por* ***la***$_{\text{fem}}$ *día* 'I'm going to study during the day' (participant 166, pre-stay, interview)

Iban a casa ***todos***$_{\text{masc}}$ ***los***$_{\text{masc}}$ *días* 'they went home every day' (participant 152, in-stay, narrative).

Whereas we originally analyzed this dataset using targetlikeness as the dependent variable (Gudmestad et al., 2019), in the current analysis, we sought to move away from using the prescriptive norm as a yardstick. For this reason, the current dependent variable was modifier gender (feminine or masculine). We coded for nine independent, fixed-effect variables that had been studied previously in research on grammatical gender marking in additional-language Spanish (see the **Supplementary Materials** for a table that lists the variables and their categories). Although these factors have been examined in prior investigations in order to better understand targetlike use, we explored their potential impact on the use of modifier gender in the current study.

Five factors pertained to characteristics of the noun. First, noun number differentiated between singular and plural nouns. Second, noun class distinguished between nouns that have biological gender (e.g., *mujer*$_{\text{fem}}$ 'woman') and those that have arbitrary gender (e.g., lápiz$_{\text{masc}}$ 'pencil'). Third, each token was coded according to the ending seen on the noun. We distinguished four categories for the variable of noun ending. Canonical *-o*/*-a* endings were masculine nouns ending in *-o* and feminine nouns ending in *-a*. Non-canonical *-o*/*-a* endings were the opposite – feminine nouns ending in *-o* and masculine nouns ending in *-a*. Predictive endings were those that, according to Teschner and Russell (1984), were strongly linked with one gender (e.g., *-ción*, as in *educación*$_{\text{fem}}$ 'education' and *-e*, as in *pie*$_{\text{masc}}$ 'foot', as feminine and masculine endings, respectively). Other endings were those that were not strongly connected to one gender (e.g., *-s* as in *país*$_{\text{masc}}$ 'country'and *tesis*$_{\text{fem}}$ 'thesis'; Teschner and Russell, 1984). The final two factors that targeted characteristics of the noun were included in order to examine the possible role that noun frequency plays in the use of modifier gender. The factor noun log-frequency (language) provided a measure of noun frequency in Spanish and, as such, is taken as a proxy for possible input. For this factor, we identified the frequency per million words with which each noun in our dataset occurred in the *Corpus del español* (Davies, 2016-). Because of the skew in the distribution of frequency scores, we used the natural logarithm of noun frequency in our analysis. The factor noun frequency (individual) provided a measure of noun frequency that depended on each individual's use. Usage-based research, which demonstrates that an individual's language use shapes her/his internal grammar (e.g., Bybee, 2006), motivates this variable. For this factor, we counted how often each participant produced a given noun with a gender-marked modifier. Because individual speakers can change their use of nouns and modifiers as a function of task and time, we calculated this score for every individual, each task, and at each data-collection period. Therefore, this coding gives the possibility of nine different frequency scores (three tasks x three data-collection periods) for a given noun for every participant. We examined the possible role that frequency plays in the use of modifier gender because previous research suggests that noun frequency influences additional-language gender marking (e.g., Sabourin et al., 2006); two different frequency factors were included because the operationalization of frequency is complex (Hashimoto and Egbert, 2019).

The remaining four variables pertained to characteristics that did not concern solely the noun. Syllable distance measured the number of syllables between the modifier and the noun. Modifier type differentiated between determiners and adjectives. For the final fixed effects, time distinguished between the pre-stay, in-stay, and post-stay data-collection periods and task analyzed possible differences among the oral interview, oral narration, and written essay. It is important to note that, while noun gender has been widely studied in investigations on grammatical gender, we have not included it in the present analysis because it can be interpreted to represent a native-speaker or prescriptive benchmark. Namely, the examination of noun gender as a fixed effect and modifier gender as the dependent variable would allow for observations about targetlike use, as the results would show, for example, whether feminine modifiers were more likely to occur with feminine nouns (i.e., targetlike use) or masculine nouns (i.e., non-targetlike use). In other words, including noun gender as a fixed effect could be considered an indirect inclusion of a native-speaker or prescriptive norm. Given the overarching goal of the current study, which is to move beyond assessments of targetlikeness or accuracy in the study of grammatical gender, we elected not to analyze this factor.

Finally, we included participant and noun type as random effects in the analysis. The participant random effect enables us to account for variability among the participants and the noun-type random effect recognizes that language behavior with individual nouns may differ. By including these two variables as random effects, we treated participants as part of a larger population of speakers and noun types as a part of a larger vocabulary. The inclusion of noun type as a random effect explains why the current dataset is limited to nouns that were used more than once, as we cannot distinguish how much variability in usage can be attributed to nouns that occur only once in the dataset.

We analyzed the data quantitatively using R (R Core Team, 2019) and SAS software, Version 9.4 of the SAS System for Windows (Copyright © 2018 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, United States). First, with R, we examined whether there were strong correlations among any of the fixed effects using chi-square methods, with the intention to remove variables from further analysis when strong correlations were observed. With bootstrapping, we also explored whether any of the fixed effects appeared to be important for explaining variation in the dependent variable. This step enabled us to remove variables that

were not important and helped us to avoid overfitting the model. Next, we fit one generalized linear mixed-effects model employing a backward selection strategy with SAS. The model was fit with respect to feminine usage. We focused on feminine modifiers because previous research has indicated that the use of feminine and masculine modifiers develops at different rates, with feminine modifiers developing more slowly (e.g., Finnemann, 1992; Montrul et al., 2008). However, it is worth clarifying that all effect estimates in this model refer to how feminine modifiers behave in relation to masculine modifiers. Thus, because the dependent variable is binary, information about the use of masculine modifiers is also present in the models. This regression examined multiple independent variables concurrently and determined which ones conditioned the use of modifier gender. If a variable was found to be non-significant, we removed it from the analysis and reran the model. For nominal independent variables, one category is selected as the reference point and compared to the other category or categories of the same variable. The reference points for the categorical fixed effects were singular (noun number), arbitrary (noun class), canonical *-o*/*-a* (noun ending), determiner (modifier type), pre-stay (time), and essay (task). Noun log-frequency (language), noun frequency (individual), and syllable distance were continuous factors, so they had no reference point. Once we identified the significant fixed effects, we examined interactions between these factors and time, in order to make observations about language development over the 21-month period covered by the LANGSNAP corpus. After fitting the model, we assessed whether any of the fixed effects were highly correlated, which would have led to instability of effect estimates. We considered a magnitude of greater than 0.6 to be highly correlated. We also identified the McFadden's R^2 (with R) and the Bayesian Information Criterion (BIC, calculated with SAS), two metrics that indicate whether the generalized linear mixed-effects model does a good job of modeling the data.

RESULTS

We began the analysis by examining whether any of the fixed effects were highly correlated. No such strong correlations were found. We then moved on to the bootstrapping phase, which revealed that noun number and noun log-frequency (language) appeared not to be important for the use of modifier gender, so we removed them from further investigation. Then we fit a generalized linear mixed-effects model. Noun frequency (individual) and noun class were not significant in this model, so we removed them and reran the model. **Table 1** shows an overview of the fixed effects included in the final regression model. The significant fixed effects were noun ending, syllable distance, task, and modifier type. Time is also included in the model, though it is not significant, because we were interested in exploring interactions between time and the fixed effects. Interactions between time and other fixed effects allow us to make observations about longitudinal development. One significant interaction was identified: Time x noun ending.

TABLE 1 | Overview of the fixed effects in the regression model.

Effect	Df	*F*	*p*
Time	2	0.41	0.6064
Noun ending	3	4.23	0.005
Syllable distance	1	9.16	0.003
Task	2	3.60	0.027
Modifier type	1	25.45	<0.0001
Noun ending x Time	6	4.439	0.0002

After fitting the model, we examined whether there were strong correlations between any of the fixed effects and found none.

Table 2 provides the details of the results for fixed effects and the interaction; the results for the two random intercepts (participant and noun type) are available as **Supplementary Materials**. We focus our presentation of the findings on three pieces of information in the tables: The estimate, *p* value, and confidence interval (CI). A positive estimate indicates a higher log-odds of using a feminine modifier compared to a masculine modifier and a negative estimate means that the log-odds of using a feminine modifier are lower. Results for the use of masculine modifiers can be inferred from these results: If we were to rerun the analysis where masculine modifiers were treated as the reference point, we would obtain effect estimates that are of the same magnitude as what we report in **Table 2**, but of opposite sign. In other words, a positive estimate for feminine modifiers would be a negative estimate for masculine modifiers and vice versa. The *p* value shows whether the result is significant (in the current analysis $\alpha = 0.05$ and significance is when $p < \alpha$.). For nominal fixed effects that have more than two categories (time, noun ending, and task), we also look to the CIs to see whether the non-reference point categories are similar to or different from each other. Overlap in CIs indicates a similarity, whereas the lack of overlap points to an important difference.

Beginning with time, although we saw that the overall *F* test is not significant (**Table 1**), the pairwise comparison of pre-stay versus in-stay was significant (**Table 2**). This indicates that collectively the variability in data-collection periods was similar, but individually two of the points were different from each other. Namely, the log-odds of using a feminine modifier were higher at in-stay than at pre-stay. There was no significant difference in the use of feminine modifiers between post-stay and pre-stay. Moreover, the overlap in the CIs for in-stay and post-stay demonstrate that the use of feminine versus masculine modifiers was similar between these two data-collection periods. For noun ending, the log-odds of using a feminine modifier were significantly higher with both non-canonical *-o*/*-a* and predictive endings compared to canonical *-o*/*-a* endings. Other endings were not significantly different from canonical *-o*/*-a* endings. The examination of the CIs of the non-reference point categories showed that the log-odds of feminine modifier use were similar to each other. Regarding syllable distance, the log-odds of using a feminine modifier decreased as the distance between the noun and the modifier increased. For task, the log-odds of feminine-modifier use were

TABLE 2 | Details of the fixed effects in the regression model.

Effect	Estimate	*SE*	Df	*t*	*p*	CI
Intercept	−1.588	0.391	20	−4.06	0.0006	[−2.404, −0.772]
Time [pre-stay]						
in-stay	0.537	0.161	10,605	3.34	0.0008	[0.222, 0.852]
post-stay	0.354	0.191	10,605	1.85	0.0643	[−0.021, 0.728]
Noun ending [canonical *-o/-a*]						
non-canonical *-o/-a*	1.699	0.798	10,605	2.13	0.0333	[0.134, 3.263]
predictive	1.504	0.583	10,605	2.58	0.0099	[0.361, 2.647]
other	−0.036	0.531	10,605	−0.07	0.9455	[−1.076, 1.004]
syllable distance	−0.058	0.019	10,605	−3.03	0.0025	[−0.096, −0.021]
Task [written]						
interview	−0.325	0.177	10,605	−1.84	0.0660	[−0.671, 0.022]
narrative	−0.662	0.248	10,605	−2.66	0.0077	[−1.149, −0.175]
Modifier type [determiner]						
adjective	−0.583	0.116	10,605	−5.04	<0.0001	[−0.809, −0.356]
Noun ending x time [canonical *-o/-a* and pre-stay]						
non-canonical *-o/-a* x in-stay	−1.831	0.417	10,605	−4.39	<0.0001	[−2.642, −1.013]
non-canonical *-o/-a* x post-stay	−1.547	0.529	10,605	−2.92	0.0035	[−2.584, −0.509]
other x in-stay	−0.143	0.278	10,605	−0.52	0.6065	[−0.688, 0.402]
other x post-stay	−0.116	0.346	10,605	−0.34	0.7375	[−0.794, 0.562]
predictive x in-stay	−0.600	3.462	10,605	−1.73	0.0832	[−1.278, 0.079]
predictive x post-stay	0.375	0.433	10,605	0.87	0.3862	[−0.474, 1.224]

Note. *The model fits the log-odds of the usage of feminine modifiers. The reference points for the nominal fixed effects and the interaction are in brackets.*

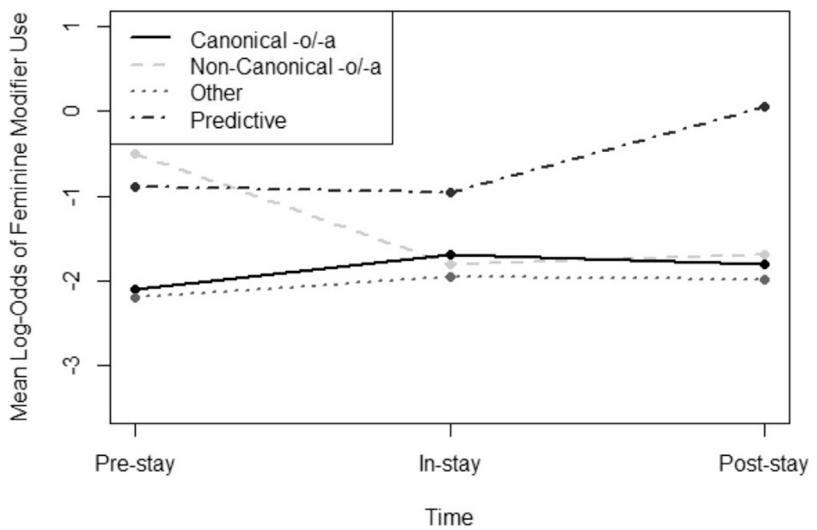

FIGURE 1 | The interaction between noun ending and time.

lower in the narrative task than in the written essay. The difference between the interview and the essay was not significant, and a comparison of the CIs for the narrative task and the interview show that the use of modifier gender was similar between the two. Additionally, the log-odds of using a feminine modifier were lower with adjectives compared to determiners (the modifier type factor).

Turning to the significant interaction, the plot in **Figure 1** illustrates the findings shown in **Table 2** for the noun ending × time interaction. A significant change over time was observed in the log-odds of feminine modifier use with non-canonical *-o/-a* nouns: The comparisons between pre-stay and in-stay and between pre-stay and post-stay show that use of feminine modifiers with such nouns (as compared to the reference points) became less likely. In sum, this participant group's variability in the use of modifier gender was conditioned by time, noun ending, syllable distance, task, modifier type, and the interaction between time and noun ending.

Finally, both the McFadden's R^2 (Smith and McKenna, 2013) and the BIC (Kass and Raftery, 1995) show that this generalized linear mixed-effects model does a good job of fitting the data for modifier gender. McFadden's R^2 is a measurement of the relative likelihood of

the fitted and null models. This metric indicates a strong model fit ($R^2_{McFadden}$ = 0.672). Next, the BIC is a metric that compared our model to a null model using their log-likelihood. It also penalizes models that have potentially extraneous parameters, which helps to protect against overfitting a model. Our model has a BIC of 5,107.73 and the null model has a BIC of 15,569.31. A difference between the two models that is larger than 10 is deemed to be strong evidence for the model with the smaller BIC. Thus, we can conclude that our model, with its lower BIC, has a higher probability of being the true model when considered against the null model.

DISCUSSION

We now return to the research question: What linguistic and extra-linguistic factors predict the variable use of modifier gender over time? We first answer this question with a discussion of the insights into the acquisition of grammatical gender marking that have emerged from the variationist analysis. We then turn to a reflection on implications of and challenges associated with SLA research that seeks to divorce itself from native-speaker and prescriptive biases.

Insights into the Use of Modifier Gender Using a Variationist Approach

Unlike previous research on the acquisition of grammatical gender marking, the present investigation did not incorporate a native-speaker or prescriptive target (cf. Alarcón, 2014). In particular, we sought to understand what factors influence the variable use of modifier gender as opposed to understanding what factors influence targetlike behavior. This methodological and conceptual decision has resulted in a new knowledge base that crucially differs from the one that has been built on a focus on accuracy. Collectively, we found that participants' variable use of modifier gender is complex and conditioned by multiple factors simultaneously. The mixed-effects model, which accounted for variability in gender-marking behavior among individual participants and noun types by including them as random effects, revealed that the use of feminine modifiers was predicted by 1) the linguistic factors of noun ending, syllable distance, and modifier type, 2) the extralinguistic factors of task and time, 3) and the interaction between noun ending and time. Among those factors, we see several indications of stability over time. Over the 21 months examined in the present investigation, participants were more likely to use feminine modifiers when little distance separated the noun from the modifier, when the modifier was a determiner (versus an adjective), and on the written essay (versus the oral narrative). Taken together, these findings show that, despite a change in learning context, there was stability over time in the factors influencing their use of modifier gender. This stability echoes research that has found that additional-language learners do not always show changes in grammar during a stay abroad (Llanes, 2011).

Stability, however, is not the whole story, as the results also showed evidence of change over time in two ways. One is that the participants were more likely to use feminine modifiers at in-stay compared to pre-stay, which points to an increase in the use of feminine modifiers over the course of an academic year in a Spanish-speaking country. Keeping in mind previous research that observed feminine-modifier use to develop more slowly than masculine-modifier use, leading researchers to suggest that the masculine modifier is the default (e.g., Montrul et al., 2008), our finding may be indicative of the fact that during an academic year abroad, the strength of the masculine default weakened as the participants became more likely to use feminine modifiers. The other evidence of development is seen with the interaction between noun ending and time: The participants were less likely to use feminine modifiers with nouns that have non-canonical *-o/-a* endings over time. In other words, the log-odds of using feminine modifiers with nouns with non-canonical *-o/-a* endings, such as *problema* 'problem', *día* 'day', *mano* 'hand', changed significantly after an academic year in Spain or Mexico, and this change was maintained after their return to the United Kingdom. Despite evidence of stability during the 21-month period, which included a stay abroad, some development in their gender-marking behavior was observed.

Thus, this variationist analysis was able to offer new details about the additional-language development of gender marking by beginning to explain the variability present in the use and development of modifier gender. In particular, we drew on three features of variationist SLA in order to address this issue. One is that the object of study was the linguistic forms (feminine and masculine modifiers) that learners used. The second characteristic was that we aimed to explain variability in modifier gender by conducting a multivariate analysis that revealed how a range of linguistic and extra-linguistic factors influenced learners' systematic variable use. Third, with the help of the LANGSNAP corpus' longitudinal data, we investigated additional-language development by exploring whether the factors that impacted the use of modifier gender changed over time. Thus, despite the assumption of consensual norms that underpins much variationist research on instances of Type II variation, our analysis offers a proof of concept for the fruitful extension of the variationist framework to analyses that remain independent of this norm.

Challenges and Implications

Before concluding we offer a reflection on some of the challenges we faced in trying to do an analysis that is independent from a native-speaker or prescriptive norm. We comment on the decisions we made in response to these challenges, which we believe may have implications for future research in SLA.

The first challenge we confronted was precisely *how* to conduct an analysis of gender marking that moved away from native-speaker and prescriptive standards for additional-language learning. Whereas scholars advocating for this paradigm shift within the field of SLA have presented convincing theoretical and conceptual arguments (e.g., Bley-Vroman, 1983; Klein, 1998; Ortega, 2016), there is still work to be done in order to work out the concrete details as to how to go about conducting this type of analysis.[11] Following a call for reform that encouraged different ways of achieving this goal (Ortega, 2017), we chose to

[11]See Murahata et al. (2016) for a methodological discussion within the multicompetence approach.

explore how an existing approach to SLA could be useful, and we opted for the variationist approach. Variationist SLA provided us with a framework in which our object of study shifted from an assessment of accuracy or targetlikeness, as was typical in previous research on gender marking in additional-language Spanish (cf. Alarcón, 2014), to an examination of the forms that participants used (i.e., feminine and masculine modifiers). Unlike traditional variationist SLA scholarship, however, we did not compare the additional-language participants in the current study to a group of native speakers of Spanish. These decisions allowed us to move both conceptually and methodologically away from native-speaker and prescriptive biases. Conceptually, we reconceived of additional-language development of gender marking as the use and evolution in the use of modifier gender, rather than improvement in accuracy. Methodologically, we introduced a new way of analyzing gender marking through a change in the dependent variable and we employed the multivariate analytical tools common in variationist SLA in order to explain the complex and systematic variability in the use of modifier gender. With variationist SLA, we were able to explain the intricacies in participants' variable use of gender marking and how this variability changed over time. In sum, we believe that the new knowledge that emerged from the current study demonstrates that the variationist approach can be beneficially adapted in order to move SLA away, both conceptually and methodologically, from native-speaker and prescriptive biases, and we believe that it is worth considering other existing frameworks to see how they might also be valuable in contributing to this paradigm shift.

A second challenge we encountered in trying to move away from a native or prescriptive norm was whether we should assess whether participants were becoming more proficient with grammatical gender (i.e., language development) and if so, how we should go about this kind of assessment (cf. Birdsong and Gertken, 2013). The majority of research within SLA makes reference to language development, with greater proficiency generally considered to correspond to language use or knowledge that is more in line with native-speaker use or knowledge or with prescriptive descriptions. In our analysis, we were able to make observations about development over time but these observations were not connected to notions of proficiency. Indeed, we saw that the use of feminine modifiers increased after a year spent in a target-language environment and that the influence of noun ending on the use of modifier gender evolved over the course of 21 months. While these insights contribute to a better understanding of the use of feminine (versus masculine) modifiers, they do not allow us – and certainly they were not designed or intended to allow us – to speak in terms of improvement per se. This is in stark contrast to most SLA research in general and to previous research on grammatical gender marking in Spanish in particular, where proficiency assessment has been done by examinations of accuracy or targetlikeness. This type of approach perpetuates the comparison of additional-language speakers to a native-speaker or prescriptive baseline. If we lose that baseline, we are necessarily confronted with the question of what is meant by proficiency and how to assess it. One possible solution, according to Cook (2016) and Ortega (2017), that would allow the field to continue to make observations in terms of proficiency would be to use other additional-language or multilingual speakers as comparison (or baseline) groups, as long as the comparison does not perpetuate a deficit view of language acquisition.[12] In other words, comparing the multilingual speakers in the current study to another group of multilingual speakers can be a way of assessing proficiency with grammatical gender marking. However, we believe that this too begs the question of what it means to be proficient even for a multilingual comparison group. More specifically, in the absence of a benchmark, how might one determine that another group of multilingual speakers is indeed more proficient than the group under study? If SLA continues to be interested in questions of proficiency and if the field moves away from native-speaker and prescriptive norms, we agree with researchers such as Piller (2002) that this presumably requires new conceptualizations of proficiency (see also Monteiro et al., 2018).

Yet another challenge we faced concerned the role that previous research, which was shaped by a native-speaker norm, should have in the current study. This came up in two ways. First, it has been common practice for research to refer to previous studies in order to identify and motivate the variables that are examined in subsequent investigations. We have followed suit with the current study. Our dependent variable of modifier gender stems from previous (target-oriented) research that has observed that additional-language participants are more accurate in marking grammatical gender with masculine nouns than feminine nouns. Furthermore, all of our independent variables had been investigated previously in work on grammatical gender (cf. Gudmestad et al., 2019). The advantage is that they are justified by past research. However, one might question this decision on at least two grounds. First, why might we believe that the same variables thought to influence targetlike use of gender marking would also be involved in explaining the use of modifier forms? Second, is not the reliance on previous (targetlike-oriented) research for the identification of variables a way of introducing native-speaker bias into a project that precisely set out to avoid such bias? With respect to this first question, we decided to examine these factors in the current study because the linguistic phenomenon under investigation was the same (i.e., grammatical gender marking), even though the object of study had shifted (i.e., modifier gender rather than targetlikeness). In other words, in the absence of previous research on modifier use to guide us, we hypothesized that factors thought to influence targetlike gender marking might also impact the variable use of modifier gender. With respect to the second question, in the current study, we considered that the potential to introduce native-speaker or prescriptive bias differed as a function of the independent variable in question. Namely, we differentiated between, on the one hand, the factors of noun gender and initial proficiency and, on the other, other factors identified in previous research. This was done precisely in order to move away from a native-speaker or prescriptive target.

[12]On the basis of the present results, another possible avenue may be to explore development from the perspective of a move away from use characterized by default forms. In this case, greater proficiency would be defined as evolution from general reliance on default forms (such as masculine modifiers) towards greater use of non-default forms (such as feminine modifiers), as was the case between pre-stay and in-stay in the current study. Care would be needed, though, to make sure that the conceptualization and analysis of default forms is clearly distinct from notions of native-speaker and prescriptive norms.

Although noun gender has been extensively studied in the previous (target-oriented) research on gender marking, prescriptive grammar rules dictate that noun gender is the sole feature that conditions the use of modifier gender and, aside from a small group of nouns whose gender differs by geographical area, native speakers do not appear to exhibit sociolinguistic variation in their use of gender marking (cf. Gudmestad et al., 2019). This suggests that native-speaker use is influenced by one factor alone – noun gender –, just as grammar rules prescribe. Thus, in order to carry out an analysis of gender marking that was not reliant on native-speaker and prescriptive biases, we decided not to include noun gender as a potential explanatory factor in our analysis because the results from this factor can be interpreted to represent a native-speaker or prescriptive norm. A similar motivation led us to exclude initial proficiency from the analysis. Although an initial proficiency score was obtained for each participant at the outset of the project using an elicited imitation task, such a score reflects the participants' ability to imitate prescriptively accurate forms, including gendered forms. For this reason, the measure of (prescriptive) initial proficiency was excluded. Thus, the current analysis consisted only of factors that could be used to characterize language use without an implicit or explicit reference to a native-speaker or prescriptive standard. By relying on previous research to design the current study, one might suggest that our strategy for reform is one that Ortega (2017) would call "modest" and a more "ambitious" strategy would be a bottom-up one in which researchers conduct detailed, qualitative analyses of multilingual language use in order to identify emergent variables.

Additionally, we debated how or whether research that has attempted to move away from a native-speaker bias should make connections with previous work that was impacted by this bias. Again, it is common practice in SLA for researchers to make connections among investigations and to be explicit about how one study builds on existing knowledge. However, moving away from assessments of language that compare additional-language participants to native-speaker and prescriptive norms is a notable conceptual change. With such a change, how do researchers succeed in making connections between differently oriented analyses in order to build new knowledge or is it worthwhile to even make these comparisons? Or, does the field of SLA need to build an entirely parallel body of knowledge? Once again, we speculated that the extensive previous research provided a relevant starting point for the present investigation, in so far as prior studies informed our selection of the dependent variable and independent variables. Only subsequent research can show whether this is a justified position on our part. Importantly, we have not made explicit comparisons between the current study's results and prior investigations because they are not on the same footing, due to the difference in the object of study (i.e., accuracy versus modifier gender). Nevertheless, we feel that this is an important issue that SLA needs to consider if this paradigm shift becomes more integrated into the field.

CONCLUSION

In the present investigation, we heeded the call for SLA research to depart from analyses that compare additional-language participants to native-speaker or prescriptive benchmarks. In so doing, we drew on strengths of an existing framework in the field – variationist SLA – in order to conduct an analysis of grammatical gender marking in Spanish that was independent from native-speaker and prescriptive targets. We shifted the object of study from accuracy to one that centered on additional-language participants' use of gender-marked modifiers. The results revealed new observations about the acquisition of gender marking in additional-language Spanish. Specifically, the variable use of modifier gender was conditioned by both linguistic (noun ending, syllable number, and modifier type) and extra-linguistic (time and task) factors, and noun ending helped to explain changes along the developmental trajectory. The current study has also demonstrated that existing approaches to SLA research can be adapted in order to help the field move beyond its native-speaker bias. Moreover, we discussed some of the challenges that we encountered when attempting to integrate this conceptual and methodological change into research on a well-studied linguistic phenomenon. In order to encourage researchers to reflect on how to respond to calls to avoid native-speaker and prescriptive biases, we feel that it is important to identify challenges inherent in this paradigm shift, as well as potential solutions to the issues encountered. By publicly reflecting on these challenges, we hope to encourage further dialogue on these issues.

DATA AVAILABILITY STATEMENT

The data presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://langsnap.soton.ac.uk/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Southhampton, United Kingdom. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AG and AE contributed to the conception and design of the study. AG is primarily responsible for the data coding. TM is largely responsible for the statistical analysis. AG wrote the paper. AG and AE contributed to revising the submitted version.

FUNDING

The publication of this paper was funded by Virginia Tech's Open Access Subvention Fund.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.723496/full#supplementary-material

REFERENCES

Alarcón, I. V. (2010). Gender Assignment and Agreement in L2 Spanish: The Effects of Morphological Marking, Animacy, and Gender. *Stud. Hispanic Lusophone Linguist.* 3, 267–300. doi:10.1515/shll-2010-1076

Alarcón, I. V. (2011). Spanish Gender Agreement under Complete and Incomplete Acquisition: Early and Late Bilinguals' Linguistic Behavior within the Noun Phrase. *Bilingualism* 14, 332–350. doi:10.1017/s1366728910000222

Alarcón, I. (2014). "Grammatical Gender in Second Language Spanish," in *The Handbook of Spanish Second Language Acquisition*. Editor K. L. Geeslin (Malden, MA: Wiley Blackwell), 202–218. doi:10.1002/9781118584347.ch12

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Birdsong, D., and Gertken, L. M. (2013). In Faint Praise of Folly: A Critical Review of Native/Non-Native Speaker Comparisons, With Examples of Native and Bilingual Processing of French Complex Syntax. *Lia* 4, 107–133. doi:10.1075/lia.4.2.01bir

Bley-Vroman, R. (1983). The Comparative Fallacy in Interlanguage Studies: the Case of Systematicity1. *Lang. Learn.* 33, 1–17. doi:10.1111/j.1467-1770.1983.tb00983.x

Bruhn de Garavito, J. B., and White, L. (2002). "The Second Language Acquisition of Spanish DPs: The Status of Grammatical Features," in *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection*. Editors A. T. Pérez-Leroux and J. M. Liceras (New York: Bilingual Press), 153–178. doi:10.1007/978-94-010-0291-2_6

Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language* 82, 711–733. doi:10.1353/lan.2006.0186

Cook, V., and Wei, L. (Editors) (2016). *The Cambridge Handbook of Linguistic Multi-Competence* (Cambridge, UK: Cambridge University Press).

Cook, V. J. (1992). Evidence for Multicompetence. *Lang. Learn.* 42, 557–591. doi:10.1111/j.1467-1770.1992.tb01044.x

Cook, V. (1997). Monolingual Bias in Second Language Acquisition Research. *Revista Canaria de Estudios Ingleses* 34, 35–49.

Cook, V. (2016). "Premises of Multi-Competence," in *The Cambridge Handbook of Linguistic Multi-Competence*. Editors V. Cook and L. Wei (Cambridge, UK: Cambridge University Press), 1–25.

Dabrowska, D. (2012). Different Speakers, Different Grammars: Individual Differences in Native Language Attainment. *Linguist. Approaches to Bilingual.* 3, 219–253. doi:10.1075/lab.2.3.01dab

Davies, M. (2016-). *Corpus del español: Two billion words, 21 countries*. Available at: http://www.corpusdelespanol.org/web-dial (Accessed July 2017).

de Prada Pérez, A. (2015). "First Person Singular Subject Expression in Spanish in Contact with Catalan," in *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective*. Editors A. M. Carvalho, R. Orozco, and N. Lapidus Shin (Washington, D. C.: Georgetown University Press), 121–142.

Dewaele, J.-M., and Véronique, D. (2001). Gender Assignment and Gender Agreement in Advanced French Interlanguage: A Cross-Sectional Study. *Bilingualism* 4, 275–297. doi:10.1017/s136672890100044x

The Douglas Fir Group (2016). A Transdisciplinary Framework for SLA in a Multilingual World. *Mod. Lang. J.* 100, 19–47. doi:10.1111/modl.12301

Fernández-García, M. (1999). "Patterns of Gender Agreement in the Speech of Second Language Learners," in Advances In Hispanic Linguistics: Papers From the 2nd Hispanic Linguistics Symposium. Editors J. Gutiérrez-Rexach and F. Martínez-Gil (Somerville, MA: Cascadilla Press), 3–15.

Finnemann, M. D. (1992). Learning Agreement in the Noun Phrase: The Strategies of Three First-Year Spanish Students. *Int. Rev. Appl. Linguistics Teach.* 30, 121–136. doi:10.1515/iral.1992.30.2.121

Geeslin, K. L., and Long, A. Y. (2014). *Sociolinguistics and Second Language Acquisition: Learning to Use Language in Context*. New York, NY: Routledge. doi:10.4324/9780203117835

Geeslin, K. L., García-Amaya, L. J., Hasler-Barker, M., Henriksen, N. C., and Killam, J. (2012). "The L2 Acquisition of Variable Perfective Past-Time Reference in Spanish in an Overseas Immersion Setting," in Selected Proceedings of the 14th Hispanic Linguistics Symposium. Editors K. Geeslin and M. Díaz-Campos (Somerville, MA: Cascadilla Proceedings Project), 197–213.

Gudmestad, A., Edmonds, A., and Metzger, T. (2019). Using Variationism and Learner Corpus Research to Investigate Grammatical Gender Marking in Additional Language Spanish. *Lang. Learn.* 69 (4), 911–942. doi:10.1111/lang.12363

Hall, J. K., Cheng, A., and Carlson, M. T. (2006). Reconceptualizing Multicompetence as a Theory of Language Knowledge. *Appl. Linguist.* 27, 220–240. doi:10.1093/applin/aml013

Hashimoto, B. J., and Egbert, J. (2019). More Than Frequency? Exploring Predictors of Word Difficulty for Second Language Learners. *Lang. Learn.* 69, 839–872. doi:10.1111/lang.12353

Kanwit, M. (2017). What We Gain by Combining Variationist and Concept-Oriented Approaches: The Case of Acquiring Spanish Future-Time Expression. *Lang. Learn.* 67, 461–498. doi:10.1111/lang.12234

Kass, R. E., and Raftery, A. E. (1995). Bayes Factors. *J. Am. Stat. Assoc.* 90 (430), 773–795. doi:10.1080/01621459.1995.10476572

Kinginger, C. (2009). *Language Learning and Study Abroad: A Critical reading of Research*. London: Palgrave Macmillan. doi:10.1057/9780230240766

Klein, W. (1998). The Contribution of Second Language Acquisition Research. *Lang. Learn.* 48, 527–549. doi:10.1111/0023-8333.00057

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Llanes, À. (2011). The many Faces of Study Abroad: An Update on the Research on L2 Gains Emerged During a Study Abroad Experience. *Int. J. Multiling.* 8, 189–215. doi:10.1080/14790718.2010.550297

Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-native Speakers*. Cambridge: Cambridge University Press.

Mitchell, R., Tracy-Ventura, N., and McManus, K. (2017). *Anglophone Students Abroad: Identity, Social Relationships, and Language Learning*. New York, NY: Routledge. doi:10.4324/9781315194851-2

Monteiro, K. R., Crossley, S. A., and Kyle, K. (2018). In Search of New Benchmarks: Using L2 Lexical Frequency and Contextual Diversity Indices to Assess Second Language Writing. *Appl. Linguist.* 41, 280–300. doi:10.1093/applin/amy056

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

Mougeon, R., Nadasdi, T., and Rehner, K. (2010). *The Sociolinguistic Competence of Immersion Students*. Clevedon, UK: Multilingual Matters. doi:10.21832/9781847692405

Mulder, K., and Hulstijn, J. H. (2011). Linguistic Skills of Adult Native Speakers, as a Function of Age and Level of Education. *Appl. Linguist.* 32, 475–494. doi:10.1093/applin/amr016

Murahata, G., Murahata, Y., and Cook, V. (2016). "Research Questions and Methodology of Multi-Competence," in *The Cambridge Handbook of Multi-Competence*. Editors V. Cook and L. Wei (Cambridge, UK: Cambridge University Press), 26–49.

Nance, C., McLeod, W., O'Rourke, B., and Dunmore, S. (2016). Identity, Accent Aim, and Motivation in Second Language Users: New Scottish Gaelic Speakers' Use of Phonetic Variation. *J. Socioling.* 20 (2), 164–191. doi:10.1111/josl.12173

O'Rourke, B., Pujolar, J., and Ramallo, F. (2015). New Speakers of Minority Languages: The Challenging Opportunity – Foreword. *Int. J. Sociol. Lang.* 231, 1–20. doi:10.1515/ijsl-2014-0029

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

Ortega, L. (2014). "Ways Forward for a Bi/multilingual Turn in SLA," in *The Multilingual Turn: Implications for SLA, TESOL, and Bilingual Education*. Editor S. May (New York: Routledge), 32–53.

Ortega, L. (2016). "Multi-competence in Second Language Acquisition: Inroads into the Mainstream?," in *The Cambridge Handbook of Multi-Competence*. Editors V. Cook and L. Wei (Cambridge, UK: Cambridge University Press), 50–76.

Ortega, L. (2017). *The Bi/multilingual Turn in SLA: How Far Have We (Not) Come, and Why? 36th Second Language Research Forum*. Columbus, OH,: The Ohio State University. October 12-15.

Picoral, A., and Carvalho, A. M. (2020). The Acquisition of Preposition + Article Contractions in L3 Portuguese Among Different L1-Speaking Learners: A Variationist Approach. *Languages* 5, 45. doi:10.3390/languages5040045

Piller, I. (2002). Passing for a Native Speaker: Identity and success in Second Language Learning. *J. Socioling.* 6, 179–208. doi:10.1111/1467-9481.00184

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Regan, V., Howard, M., and Lemée, I. (2009). *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*. Clevedon, UK: Multilingual Matters. doi:10.21832/9781847691583

Rehner, K. (2002). The Development of Aspects of Linguistic and Discourse Competence by Advanced Second Language Learners of French. *Dissertation Abstr. Int.* 63, 12, 2002 . Doctoral dissertation.

Sabourin, L., Stowe, L. A., and de Haan, G. J. (2006). Transfer Effects in Learning a Second Language Grammatical Gender System. *Second Lang. Res.* 22, 1–29. doi:10.1191/0267658306sr259oa

Smith, T. J., and McKenna, C. M. (2013). A Comparison of Logistic Regression Pseudo R^2 Indices. *Mult. Linear Regression Viewpoints* 39, 17–26.

Tarone, E. (2007). Sociolinguistic Approaches to Second Language Acquisition Research-1997-2007. *Mod. Lang. J.* 91, 837–848. doi:10.1111/j.1540-4781.2007.00672.x

Teschner, R. V., and Russell, W. M. (1984). The Gender Patterns of Spanish Nouns: An Inverse Dictionary-Based Analysis. *Hispanic Linguist.* 1, 115–132.

Tracy-Ventura, N., and Huensch, A. (2018). "Chapter 6. The Potential of Publicly Shared Longitudinal Learner Corpora in SLA Research," in *Critical Reflections on Data in Second Language Acquisition*. Editors A. Gudmestad and A. Edmonds (Amsterdam: John Benjamins), 149–170. doi:10.1075/lllt.51.07tra

White, L., Valenzuela, E., Kozlowska–Macgregor, M., and Leung, Y.-K. I. (2004). Gender and Number Agreement in Nonnative Spanish. *Appl. Psycholing* 25, 105–133. doi:10.1017/s0142716404001067

Young, R. (1991). *Variation in Interlanguage Morphology*. New York, NY: Peter Lang.

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 16 September 2021
doi: 10.3389/fpsyg.2021.717379

The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry

Anna-Lena Scherger[1]*, *Gianna Urbanczik*[1], *Timon Ludwigs*[1] and *Jasmin M. Kizilirmak*[2,3]

[1] Department of Rehabilitation Sciences, Language & Communication, TU Dortmund University, Dortmund, Germany, [2] University of Hildesheim, Institute for Psychology, Neurodidactics & NeuroLab, Hildesheim, Germany, [3] German Center for Neurodegenerative Diseases, Cognitive Geriatric Psychiatry, Göttingen, Germany

The present pilot study investigated potential effects of early and late child bilingualism in highly proficient adult bilinguals. It has been shown that some early second language (eL2) speakers stagnate when it comes to complex linguistic phenomena and that they display subtle difficulties in adulthood. Therefore, we have chosen the complex structure of double object constructions. We investigate the long-term achievement in a combined-method approach using elicited production, explicit comprehension by sentence-picture matching and a measure of implicit linguistic knowledge, namely pupillometry. This eye tracking method is suitable for measuring implicit reactions of the pupils to unexpected or ungrammatical stimuli. For production, ditransitive structures were elicited by means of a game. For comprehension, a sentence-picture matching task was conducted. Two pictures were shown on a monitor that were equal with respect to the involved objects, but the thematic roles of direct and indirect objects were interchanged. Items were controlled for length, gender, animacy, semantic likelihood and word order. Reaction times and accuracy scores were analyzed. To this end, $N = 18$ bilingual adult speakers of German (+ another language, mean age: 26.5) with different ages of onset participated in this study and were compared to $N = 26$ monolingual German adult speakers (mean age 23.9). All participants had a proficiency of German above 89% correct in placement and cloze tests. Results show fully comparable productive and comprehensive competencies in monolinguals and bilinguals including the reaction times in the sentence-picture matching task and a word order effect on the reaction times in both groups. In the pupillometry task, we found monolinguals and bilinguals to be sensitive to differing conditions with respect to grammatical and ungrammatical utterances. However, we find between group differences in pupil dilations in that bilinguals react differently to strong grammatical violations than monolinguals. These results are discussed with respect to the term of native speaker competence and the variation within both groups.

Keywords: language production, language comprehension, ditransitives, implicit and explicit knowledge, article omission, pupillometry

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez, University of the Balearic Islands, Spain

Reviewed by:
Jens Schmidtke, German-Jordanian University, Jordan
Marta Velnic, Norwegian University of Science and Technology, Norway

***Correspondence:**
Anna-Lena Scherger
anna-lena.scherger@tu-dortmund.de

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

Received: *30 May 2021*
Accepted: *13 August 2021*
Published: *16 September 2021*

Citation:
Scherger A-L, Urbanczik G, Ludwigs T and Kizilirmak JM (2021) The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry. Front. Psychol. 12:717379. doi: 10.3389/fpsyg.2021.717379

INTRODUCTION

In research on bilingual language acquisition and its outcome in adulthood, it has been shown that some of the children acquiring an early second language (eL2) are doing well on this acquisition task and others struggle with more complex structures at some point in their acquisition process and stagnate or fossilize (Paradis, 2016). Paradis (2019) discusses eL2 children's outcome and ultimate attainment in adulthood. While they are able to reach high proficiency, are typically indistinguishable from monolingual speakers in conversation, and score within the normal range of monolingual performance in most language tasks, there are subtle differences between monolinguals and bilinguals with respect to production and grammaticality judgement tasks regarding complex morphosyntax in adulthood. Such long-term achievement in (e)L2 acquisition is marked by high inter-individual variability which in turn has been attributed to age, age of onset (AOO), a broad array of external environmental and social factors (such as socio-economic status or input quality and quantity), and also to individual differences in the affective and cognitive domains influencing acquisition (Granena, 2014).

In the present pilot study, we investigated in depth highly proficient bilingual adults with various ages of onset instead of focusing on moderating factors of the ability to become highly proficient in a second language (L2). For this purpose, we performed a study with respect to the different modalities of production and comprehension and implicit grammaticality judgements, regarding ditransitive structures, a complex morphosyntactic phenomenon. Our aim was to evaluate possible subtle differences between bilingual and monolingual speakers, as Paradis (2019) postulated. The purpose of this in-depth investigation of one linguistic phenomenon was to illuminate the term "native speaker" and to evaluate whether bilinguals exhibit a level of proficiency comparable to monolinguals' performance on production, comprehension and implicit grammaticality judgement tasks.

In the literature on the definition of a native speaker, we find different approaches and different acquisition types discussed (see, e.g., the discussion on native vs. non-native foreign language learners in Clahsen and Felser, 2006) as well as varying age spans investigated. Several suggestions have been made on how to distinguish between native and non-native speakers. Amongst others, Halliday (1975) characterized a native speaker as *someone who is able to predict what the other person is going to say*, therefore enabling a native speaker to anticipate upcoming input. Others suggested to characterize native speakers by their *competence to identify ungrammatical utterances* and thus to provide valid grammaticality judgements on their language using intuitive knowledge of the grammatical sentence (Chomsky, 2002).

While there seems to be a common ground on what native speakerism means, it is in many cases unclear what it does not include. It is for instance obvious that a "bilingual is not two monolinguals in one person" (Grosjean, 1989). There is cross-linguistic influence at play in bilingual development (Müller and Hulk, 2001) that is attested in many morphosyntactic domains in 2L1 and eL2 (e.g., Schmitz et al., 2012; Scherger, 2016; for a meta-analysis see van Dijk et al., 2021). This interaction of languages within a bilingual individual is "part and parcel of bilingual development" (van Dijk et al., 2021, p. 1). However, it is unclear whether early bi- or multilingual speakers are native speakers of two or more languages (Wilkinson, 2020, p. 285). Only AOO appears to be central to the definition of native speakerism: a native speaker has acquired his or her native language from birth or at least from an early age (Davies, 2004). Nonetheless, acquiring a language at an early age is no guarantee for high proficiency, and, conversely, acquiring a language during adulthood does not rule out the attainment of high proficiency (Ortega, 2019).

Beyond that, the native speaker concept gives the impression to have "a strong monolingual bias" (Dewaele, 2017, p. 236), which results in a number of issues. Firstly, the concept has been taken to imply that monolingualism is "the gold standard" (Mauranen, 2012, p. 4), which is unsurprisingly controversial. More importantly, monolingualism is rather untypical in a predominantly multilingual world, yet monolinguals are commonly used as control groups in studies on bi- and multilingualism, reinforcing the native vs. non-native dichotomy (Dewaele, 2017). From a psycholinguistic perspective, the comparison between L1 and L2 speakers, or native and non-native speakers, respectively, is a neutral, non-judgmental comparison with the aim of gaining further understanding of how language is stored, represented and processed in the human brain.

The central issue, however, appears to lie in the term "native" itself. It is commonly perceived as a synonym for high linguistic competence or proficiency, whereas non-native speakers are often considered less competent (for a discussion see Cook, 2008 and Dewaele, 2017). It is, on the other hand, commonly accepted that there is large variation within non-native but also among native speakers when it comes to proficiency. This study, therefore, aims at investigating whether complex morphosyntactic phenomena, such as ditransitive structures, reveal a dichotomy in so-called native and highly proficient non-native speakers' productive and receptive language skills. To test this, we performed experiments using both online and offline methods, i.e., tasks that investigate language during and shortly after processing. Results are intended to shed light on the question whether non-native speakers inevitably perform below the level of native speakers in tasks involving complex grammatical structures, or whether high language proficiency, albeit non-native, is on the same level as native speakers' performance.

Turning to possible stagnation/fossilization in complex morphosyntactic phenomena in eL2 acquisition, the findings by Paradis (2016, 2019) put to question whether the phenomenon under investigation is fully mastered in bilingual acquisition. In monolingual German acquisition, ditransitive structures are the latest to be acquired regarding dative case marking (Scherger, 2015) and case marking is considered a very late acquisition phenomenon in both, monolingual and bilingual acquisition (Schulz and Grimm, 2019). To the authors' knowledge, it has not been investigated so far whether a complete mastery in

eL2 bilinguals' adult outcome can be documented. Therefore, an in-depth scrutiny of ditransitives is promoted here.

Since we include highly proficient bilingual adults in our sample who had different AOOs, in what follows, we briefly present the state of research with respect to the sensitive period in bilingual language acquisition and report what previous research states about potential outcomes in adulthood. Furthermore, we introduce the linguistic phenomenon under investigation and give some background information about predictive processing and on pupillometry, which may be unknown to parts of the readership.

Sensitive Period in Bilingual Language Acquisition

A classic topic in L2 acquisition concerns the role of AOO in achieving native-like ultimate attainment. Lenneberg (1967) put forward the controversial hypothesis of a "sensitive period," suggesting that the grammar-learning ability in acquisition declines at some point. Since then, research has provided evidence that child L2 learners outperform adult L2 learners (e.g., Johnson and Newport, 1989). However, recent findings about child L2 learners have also shown that they do not always converge fully with native speakers (Paradis, 2019). AOO has been claimed to be the factor affecting such "*near*-native (rather than fully nativelike) attainment" (Bylund et al., 2021, p. 18); the maturation of the brain is one possible explanation for this. Studies report that some of the early learners and, in fact, every late learner perform "only" near-native-like when scrutinized in detail (Abrahamsson and Hyltenstam, 2009; Stölten et al., 2014). It was concluded from these differences between L1 and L2 that even short delays in language exposure (as in eL2 children) may have minor consequences for ultimate attainment (Bylund et al., 2021).

A further controversial issue is the exact offset of the sensitive period. In some studies, it turned out to be as early as 3;6 years (Meisel, 2018), and in others, it was found to be much later than previously assumed. Hartshorne et al. (2018) found in more than 600,000 children that the grammar-learning ability changes with age but is preserved almost into adulthood. They determined the age of 17;4 as the turning point from which this ability starts declining. With respect to our participants, this would mean that every bilingual participant with an AOO prior to 17;4 years is able to achieve native-like ultimate attainment.

Ditransitive Constructions and Case Marking in German

In German, ditransitives are characterized by verbs, such as *geben* (to give) or *zeigen* (to show), that select a nominative subject (SUBJ), a direct object (DO) typically marked for accusative (ACC) and an indirect object (IO) typically marked for dative (DAT). Three thematic roles are assigned (Primus, 2012): Agent to the nominative, theme to the DO (see Example 1a) and recipient to the IO (see Example 1a). Thus, grammatical information is encoded morphologically, mostly on the determiner, which allows for a relatively free word order.

(1) *a.* *ich* *gebe* *dem kind* *den ball*
i give (to) the child IO-DAT the ball DO-ACC
"i give the ball to the child"

b. *ich* *gebe* *den ball* *dem kind*
i give the ball DO-ACC (to) the child IO-DAT
"i give the ball to the child"

Example 1a illustrates the unmarked default word order IO-DO (Drenhaus, 2004; Kholodova and Allen, in press). While there are many instances, where DO-IO is the default word order (Müller, 1999) and may even be the underlying word order (Røreng, 2011), IO-DO is considered the unmarked word order for non-pronominal, full determiner phrases (DPs) when the IO is [+animate] and the DO is [-animate]. This word order corresponds to the canonical structure of *recipient-theme* proposed by Kholodova and Allen (in press), as recipients are typically [+animate]. However, when the IO is [-animate] and the DO is [+animate] the unmarked word order is DO-IO (Müller, 1999). When both objects are [+animate], IO-DO is considered the unmarked word order (Müller, 1999). The marked order for [+animate] IOs and [-animate] DOs is illustrated in Example 1b. This word order is grammatical, but rather uncommon in German.

Sauerman and Höhle (2018) showed that in child-directed speech, IO-DO is the most frequent order. As factors influencing the word order, they documented (i) animacy (animate vs. inanimate), (ii) definiteness (definite vs. indefinite), (iii) givenness (given vs. new), and (iv) reference expression (pronoun vs. full lexical phrase; Sauerman and Höhle, 2018). Furthermore, Drenhaus (2004) found effects of word order on the correct case marking in monolingual children from age 3;9 to 6;8. Word order and case marking were correctly repeated in sentences with the default word order. However, the non-default DO-IO could not be repeated correctly.

In first language (L1) and eL2 acquisition literature, there is consensus that, owing to its high complexity, German case marking is acquired very late compared to other acquisition phenomena (Schulz and Grimm, 2019). However, the exact age of mastery of this complex morphosyntactic phenomenon in different types of acquisition (Scherger, 2016, 2018; Ulrich et al., 2016; Lemmer, 2018; Schulz and Grimm, 2019) and possible fossilization of eL2 acquisition (Scherger, 2019) is still under debate. In monolingual acquisition, the reported age of mastery of the case marking paradigm is between 4;6 (Schmitz, 2006) and 9;0 years (Ulrich et al., 2016). For bilingual children, we find early primary school age as the age of mastery for simultaneous bilingual (2L1) children (Scherger, 2016), but for eL2 children, till date, no possible complete acquisition of the case paradigm has been reported. Regarding dative case marking, we know that language acquisition in eL2 children is delayed compared with L1 and 2L1 children (Lemmer, 2018; Scherger, 2019). Furthermore, prepositional case marking seems to be the case subtype acquired earliest in L1 and eL2 children (Lemmer, 2018; Scherger, 2021). Dative case marking in ditransitive structures has been shown to be the most difficult to acquire in monolingual and bilingual children's production (Scherger, 2015); therefore, these are supposed to be the most complex case structures in German matrix clauses. This is why we focus on double-object constructions with ditransitive verbs in this paper.

Before achieving target-like utterances like in Example 1, L1 children show an extended period of difficulties with these structures (Eisenbeiss et al., 2006; Schönenberger et al., 2011). Typically, nominative is acquired first, followed by accusative and dative (Eisenbeiss, 2002). Error patterns are article omission (see Example 2a by Scherger, 2015) and overgeneralization of the accusative in dative contexts (see Example 2b by Scherger, 2015). In Example (2b), there is syncretism as the definite article of feminine singular nouns is identical for nominative and accusative forms (both *die*). Here, the production of the IO *die schnecke* is assumed to be an overgeneralization of the accusative.

(2) *a.* **ich schenke hund auto*
i give (to the) dog IO-NULL (the) car DO-NULL
"i give the car to the dog"

b. **die krone kann ich die schnecke schenken*
the crown DO-ACC can i (to) the snail IO-ACC give
"i can donate the crown to the snail"

Regarding ditransitive structures, there are even fewer studies on the acquisition of comprehension capacities than on production. A related study, though not on ditransitives, was conducted by Dittmar et al. (2008). They investigated whether 2-, 5-, and 7-year-old German children are able to use the grammatical cues of case marking and word order in transitive constructions to identify agents and patients (see example 3; Dittmar et al., 2008, p. 1155).

(3) *den hund beißt der mann*
the dog DO-ACC bites the man SUBJ-NOM
"the man bites the dog"

The results showed that younger children relied predominately on word order to interpret the sentences (i.e., the dog bites the man), whereas seven-year-old children behaved like adults by relying on case markers rather than word order (i.e., the man bites the dog). In line with these results, Brandt et al. (2016) report in a pointing study that older children (aged six) relied on case marking, whereas 3- and 4-year-old children employed only word order to interpret simple transitive sentences.

In an experiment using the same methods as the present study, Scherger et al. (submitted)[1] investigated 5–7 year-old L1 children with regard to production and comprehension of ditransitives. They found that only 56.0% of these monolingual children could produce and 62.5% could comprehend as accurately as L1 adults. The results of the reaction times (RTs) in the comprehension task showed that children by that age did not react explicitly by button press in the sentence-picture matching task before hearing the second object. Monolingual adults on the other hand reacted even before the ditransitive structure's second object was auditorily presented. This was interpreted as an explicit sign of predictive processing.

Predictive Sentence Processing in L1 and L2

An implicit reflection of language processing is the prediction of upcoming input. Earlier research on predictive processing has shown that the human parser of adult native speakers predicts upcoming input before it is heard. In this context, prediction is defined as "pre-activation/retrieval of linguistic input before it is encountered by the language comprehender" (Huettig, 2015, p. 122). For instance, in an eye tracking study Altmann and Kamide (1999, p. 250) showed that on hearing a sentence like –"The boy will eat the cake," adult participants looked at the edible object out of the four objects presented already after hearing the verb "to eat," thus showing anticipatory eye movements. Instead, on hearing a sentence like "The boy will move the cake," there were no anticipations found in that there were no saccades to the cake on hearing the verb *move*. The same anticipation mechanism was found with respect to the morphosyntactic cue of case marking. Adult native speakers of German anticipated the second nominal phrase (NP) of a sentence on hearing the case-marked first NP (see Example 4; Kamide et al., 2003, p. 41).

(4) a. *der hase frisst gleich den kohl*
the hare NOM eats shortly the cabbage ACC
"the hare will shortly eat the cabbage"

b. *den hasen frisst gleich der fuchs*
the hare ACC eats shortly the fox NOM
"the fox will shortly eat the hare"

However, till date, research on adult L2 learners has shown mixed results (Kaan, 2014). Regarding transitive structures, Hopp (2015) found adult late L2 speakers in contrast to L1 speakers not to be able to use the case marking cue to predict upcoming input. Further evidence of late L2 learners not showing predictive processing (e.g., Martin et al., 2013) resulted in the hypothesis of Reduced Ability to Generate Expectations (RAGE) for late L2 learners (Grüter et al., 2014, 2017). Recently, the RAGE hypothesis was put to test for highly proficient Russian late L2 learners of German (Schlenter, 2019), investigating ditransitive structures like in example 5 (Schlenter, 2019, p. 120) in an eye tracking paradigm.

(5) a. *der gärtner gibt der blühenden pflanze eilig frisches wasser*
the gardener gives (to) the flowering plant DAT quickly fresh water ACC
"the gardener quickly gives fresh water to the flowering plant"

b. *der gärtner gibt die blühende pflanze eilig dem postboten*
the gardener gives the flowering plant ACC quickly (to) the postman DAT
"the gardener quickly gives the flowering plant to the postman"

Schlenter (2019) showed that adult speakers (L1 as well as late L2) could predict the second object through anticipatory looks before hearing it. However, note that Russian has a morphologically rich case-marking paradigm that is in many ways comparable with German. Therefore, the questions of whether Schlenter's findings can be confirmed with other L1s than Russian and whether the findings remain stable when animacy is more narrowly controlled still remain unanswered.

Additionally, a recent study on anticipatory mechanisms in early bilingual language processing by Desideri and Bonifacci (2018) even found a bilingual advantage in anticipating upcoming input. They investigated bilingual Italian-German adults from Merano, where both Italian and German are official languages. Inclusion criteria required participants to have bilingual parents and an AOO before school. Therefore, we assume these speakers to be 2L1 or eL2 bilinguals. In the experiment, participants saw four pictures on a screen and had to complete an Italian sentence such as "The woman will spread the butter on the…" by pressing the key corresponding to the matching picture (e.g., flower or bread). Bilinguals were found to outperform Italian monolinguals in this sentence completion task by showing faster RTs in choosing the target word.

Pupillometry

Pupillometry is the study of changes in pupil diameter. They are "small-scaled, rapid fluctuations in pupil diameter […] difficult to detect by unaided observation" (Beatty and Lucero-Wagoner, 2000, p. 143). However, with the advent of automatic eye trackers, which measure pupil size between 30 and 1,000 times per second, this observation has become possible. Besides the explicit responses to language described above, we are also interested in implicit signs of reactions to language input. Therefore, we use pupillometry as this method does not require explicit behavior but still identifies responses to language and acts like "a window to the preconscious" (Laeng et al., 2012).

The pupil's diameter is inherently variable with a typical size of ∼3–4 mm and a range from 1 to 9 mm (Beatty and Lucero-Wagoner, 2000; Sirois and Brisson, 2014). The pupil reacts by constriction or dilation to stimulations in ∼200 ms (Davson, 1972; Sirois and Brisson, 2014), e.g., to stimulations like varying luminance levels (Hepach and Westermann, 2016). More importantly, also other elements can influence the pupil size. Since the 1960s, research has repeatedly shown that factors such as arousal (Bradshaw, 1967; Bradley et al., 2008), emotion (Partala and Surakka, 2003; Zheng et al., 2014), attention (Karatekin, 2004), memory (Kahneman and Beatty, 1966; Papesh et al., 2012; Johnson et al., 2014), cognitive load/intensity and mental effort (Beatty, 1982; Porter et al., 2007; Piquado et al., 2010), novelty (Naber et al., 2013), and task complexity (Schluroff, 1982; Kosch et al., 2018) influence the pupil size, without the participants' knowledge (Laeng et al., 2012; Sirois and Brisson, 2014; Schmidtke, 2018; Zekveld et al., 2018).

As early as Hess and Polt (1964) measured pupil size in relation to simple multiplication tasks. Adult participants mentally calculated the product of small numbers in four tasks of varying difficulty (e.g., $7\times8 = ?$, $8\times13 = ?$, $13\times14 = ?$, $16\times23 = ?$), while changes in their pupil sizes were measured. The participants' pupil diameters increased with the increased difficulty of the calculation, indicating that task-evoked responses in the pupils are an effective way to measure processing load/effort (Beatty, 1982). Another paradigm that also focused on the brain's ability to create predictions or expectations is the so-called violation of expectation (VoE) paradigm. Here, pupillometry is used to identify surprise. When the participants' expectations of something are violated, pupils dilate (such as a violation of an expectation of a specific expected rhyme pattern; see, among others, Satterthwaite et al., 2007; Preuschoff et al., 2011; Yu, 2012; Scheepers et al., 2013; Lawson et al., 2017; Renner and Włodarczak, 2017).

With respect to linguistics, literature reports the use of pupillometry mostly in adults (Just and Carpenter, 1993; Scheepers and Crocker, 2004; Engelhardt et al., 2010; Fernandez et al., 2018; for a review, see Schmidtke, 2018), but to some extent also in infants and toddlers (Tamási et al., 2017; Süss et al., 2018). Scheepers and Crocker (2004) investigated syntactic priming with respect to case marking of subjects and direct objects regarding the subject-first preference in native German young adults (also called N1 bias, Lidzba et al., 2013). Scheepers and Crocker (2004) used pupillometry to identify garden-path effects employing abiguous structures. They report that structures that were disambiguated toward object-initial reading were harder to process than those disambiguated toward subject-initial reading. Pupil sizes have been found to increase with processing difficulty. Moreover, in a word recognition paradigm, Schmidtke (2014) investigated adult English monolinguals as well as Spanish-English early and late bilinguals (early bilinguals had an AOO before age 8 and late bilinguals had an AOO of 18 or later) with respect to their word retrieval effort. Pupil size was recorded while hearing an English word and matching it to one out of four pictures. Bilingual speakers displayed an overall delayed pupil response compared to monolinguals. Within the bilingual group, higher English proficiency was linked to an earlier response of the pupil. Thus, pupillometry was able to identify implicit word retrieval effort that differed between monolinguals and bilinguals.

With respect to VoE, pupillometry has been shown to be a useful marker for the ability of 30 months-old toddlers to differentiate between correct and incorrect pronunciations (Tamási et al., 2017), as well as 30- to 36-months-old children's sensitivity to attributive gender marking (Süss et al., 2018). As a sign of VoE, they demonstrated bigger pupil dilations in response to ungrammatical (**da ist ein blauer Haus*, 'there is a blue$_{MASC}$ house$_{NEUTR}$') than to grammatical utterances (*da ist ein blaues Haus*, 'there is a blue$_{NEUTR}$ house$_{NEUTR}$').

In sum, the core findings of pupillometry-based studies on language are that it is a valuable and valid measure of linguistic complexity. The more complex a linguistic structure, the more difficult it is to process. The more surprising an upcoming input structure is, the higher the difficulty to process, and the higher the proficiency of a speaker, the less effortful the processing of particular structures. With all of these methodological advantages regarding pupillometry in mind, we tested monolingual and bilingual adults' implicit sensitivity to grammatical violation.

Research Questions and Hypotheses

The present study investigates the questions of whether highly proficient bilingual speakers of German with various L1s reach comparable performance levels like monolinguals. Specifically, we want to answer the following questions:

Do bilingual speakers

(1) produce comparable ditransitive structures to monolinguals (production),

(2) comprehend ditransitive structures accurately (comprehension),
(3) use the case marking cue of the first object for anticipating the thematic role of the second object (predictive processing) and
(4) react implicitly to ungrammatical auditory stimuli by a change in pupil size (implicit sensitivity to grammatical violations)?

Considering prior findings, we assume that the participating L1 and 2L1 speakers can produce and understand the tested ditransitives. Therefore, we expect performance at ceiling in the production and comprehension tasks. However, we expect subtle difficulties in eL2 and late L2 learners of German regarding production of the dative case marking, because prolonged difficulties in children have been shown in Lemmer (2018) and Scherger (2019). As comprehension precedes production in acquisition (Scherger et al., submitted)[1], we do not expect any of the investigated (highly proficient) speakers to show difficulties with comprehension accuracy. However, as difficulties in anticipatory processing within late L2 learners (Grüter et al., 2014; Hopp, 2015) and fossilizations within complex morphosyntactic domains have been reported (Paradis, 2016), we expect the eL2 and late L2 speakers to show slower RTs than 2L1 and L1 speakers. Based on findings of bilinguals outperforming monolinguals in anticipating upcoming input by Desideri and Bonifacci (2018), we even expect faster RTs in the sentence-picture matching task for 2L1 bilinguals than for L1 speakers. With respect to the implicit sensitivity to grammatical violation, we expect monolingual and bilingual speakers to behave alike. We assume that the highly proficient participants of the present study should have built grammatical representations of the target structure and therefore show comparable pupil responses, when their expectations of grammatical structures are violated. We therefore do not expect differences between monolingual and bilingual participants in the implicit grammaticality judgement task.

METHODS

Participants

In this study, 44 adult speakers participated (see **Table 1**). Monolingual ($N = 26$) and bilingual ($N = 18$) participants were matched based on age (Mann-Whitney U: $U = 511.0$, $p = 0.075$). All participants were university students and received course credits for their participation in the experiments.

Monolingualism was defined as having acquired only one language (i.e., German) in the first years of life. This does of course not exclude the learning of an L2 at school. In the bilingual group, simultaneous and sequential bilinguals were included on purpose so that different AOOs could be compared (2L1: $N = 6$, AOO = 0;0; eL2: $N = 7$, AOO = 3;0–4;0; late L2 learners: $N = 5$, AOO = 6;0–13;0). Various L1s were included in the bilingual group: Albanian ($N = 2$), Arabic ($N = 1$), English ($N = 1$), Italian ($N = 1$), Kurdish ($N = 1$), Polish ($N = 4$), Russian ($N = 2$), Serbian ($N = 1$), Spanish ($N = 3$), Turkish ($N = 1$), and Vietnamese ($N = 1$). Participants were assigned to the corresponding groups according to our definition of mono- or bilingualism based on self-reports.

To evaluate the participants' German proficiency, we conducted a placement test with multiple-choice questions. Alternatives were manipulated regarding syntactic, lexical, semantic and pragmatic knowledge of German. Additionally, the participants completed two cloze tests, in which lexical, semantical and orthographical knowledge was tested. These were taken from the International study center of the University of Kassel, Germany. Each text contained about 70 words and was truncated canonically, i.e., starting with the second sentence, every second word's second half was truncated. For further evaluation of the participants' proficiency, three one-minute verbal-fluency tasks were carried out (Friesen et al., 2015; Lemmerth and Hopp, 2018). In the category-fluency task, participants were asked to name "animals" and "objects at home." In a letter-fluency task, they were asked to name words starting with the letter "s." Furthermore, we screened WM components by assessing FW and BW digit spans (WISC-V; Wechsler, 2017).

Table 2 summarizes the German proficiency and shows that no significant differences existed between monolingual and bilingual speakers. Additionally, **Table 2** lists the WM results (digit spans), which also show no group differences.

General Procedure

Participants were tested individually in a quiet room, after providing informed consent. For the comprehension task and the pupillometry experiment, viewing distance and head position were held constantly at 70 cm by a forehead and chin rest. As prior literature reported priming effects from comprehension on production but no priming effects from production on comprehension (Kauschke and Siegmüller, 2010), we conducted the production prior to the comprehension task. Finally, we conducted the additional pupillometry experiment. Overall, a test session lasted about 60 min.

STUDY 1 – PRODUCTION

Production Study Design

For eliciting ditransitive constructions, we asked the participants to play a card game with three stuffed animals (each belonging to a separate German gender: *der hund*$_{\text{MASC}}$ [the dog], *die schnecke*$_{\text{FEM}}$ [the snail] and *das schaf*$_{\text{NEUTR}}$ [the sheep]). Overall, the game consisted of 27 cards with pictures of animals. The participants had to give/donate the animals on the picture cards to one recipient (one of the three stuffed animals) describing their action. This resulted in utterances like *ich gebe das pferd dem schaf* or *ich gebe dem schaf das pferd* ("I give the horse to the sheep," see **Supplementary Table 1** for further details), where both the DO and the IO were [+animate]. There were two practice examples in which the experimenter picked a card and gave it to a stuffed animal producing an example utterance. The experimenter used IO-DO in his/her examples.

[1]Scherger, A.-L., Kizilirmak, J. M., and Folta-Schoofs, K. (submitted). Ditransitive structures in child language acquisition: an investigation of production and comprehension in children aged five to seven. *J. Child Lang.*

TABLE 1 | Participants' characteristics.

Monolinguals (*N* = 26)			
Proficiency level	<90% correct in placement test and cloze test (overall mean = 97.0%, *SD* = 2.2%)		
	Mean	**SD**	**Range**
Age	23.9	7.6	19–48
Bilinguals (*N* = 18)			
Proficiency level	<89% correct in placement test and cloze test (overall mean = 96.4%, *SD* = 2.4%)		
	Mean	**SD**	**Range**
Age	26.5	6.14	19–38
AOO German	4;1	4.14	0;0–13
LoE in years	22.6	4.9	16.5–33

AOO, age of onset; LoE, level of education.

TABLE 2 | Participants' German proficiency.

	Placement test	**Cloze test**	**Letter fluency test**	**Animals**	**Objects at home**	**Overall fluency**	**Digit span FW**	**Digit span BW**
Monolinguals								
Mean (%)	97.3	96.8	17.1	25.9	27.7	70.7	6.5	6.1
SD (%)	2.2	3.4	4.8	6.2	6.8	17.8	1.2	1.2
Range (%)	90–100	92–100	10–26	13–39	14–39	33–95	5–10	4–8
Bilinguals								
Mean (%)	97.0	95.7	20.3	26.8	29.1	76.2	6.3	5.3
SD (%)	2.7	4.0	5.4	6.6	6.4	15.9	1.1	1.4
Range (%)	92–100	87–100	10–32	15–46	18–47	55–125	3–8	3–8
Mann-Whitney U test	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$

FW, forward; BW, backward.

To exclude confounding factors, we controlled for animacy and included only [+animate] direct and indirect objects (see Gamper, 2019, for case-animacy coalitions), all of which were animals to exclude influence of animacy hierarchy. To avoid the participants' use of semantic cues for the assignment of thematic roles, all items used were semantically reversible. In sentences like those given in Example 1, the decoding of case markings is not necessary, since semantic cues assure the assignment of thematic roles (the child cannot be given to the ball, but vice versa). Furthermore, definiteness was controlled for in the production task by using only definite nouns in full lexical DPs in the practice examples given by the experimenter. This was supposed to make the participants avoid indefinite articles and pronouns in their own productions as well. Finally, to avoid verb bias, we included two verbs (*jemandem etwas geben* and *jemandem etwas schenken,* "to give something to somebody"). These two verbs are comparable with respect to their semantics, lengths, and subcategorization frames. Both are frequent and attested to be acquired early in German (Grimm and Doil, 2006).

Production Data Analysis

For the analysis of the production task, target-like accusative and dative case markings were counted separately (raw scores and percentages for each). Moreover, utterances were analyzed separately with respect to the word order produced: DO-IO (see Example 6a) or IO-DO (see Example 6b).

(6) a. *ich schenke das schaf dem hund*
i give the sheep DO-ACC (to) the dog IO-DAT
"i give the sheep to the dog"

b. *ich schenke der schnecke das pferd*
i give (to) the snake IO-DAT the horse DO-ACC
"i give the horse to the snake"

The amount of analyzed utterances in bilinguals was $N = 462$ and in monolinguals was $N = 644$ (total amount: $N = 1{,}106$).

On average, 1.9% ($SD = 1.0$) and 1.8% ($SD = 0.8$) of all the produced objects were built with pronouns by bilinguals and monolinguals, respectively. These occurrences were excluded from the analysis for a better match within the comparison of comprehension and production data. Furthermore, we excluded realizations of indirect objects by prepositional phrases (PPs), since this structure does not mandatorily require a dative case marking (see Example 7). Utterance dropout rates because of producing PPs were 3.3% (15/462) in bilinguals and 2.6% (17/644) in monolinguals. These numbers confirm results reported by Kholodova and Allen (in press) that participants

more often realized double objects as DP-DP than as DP-PP structures. Since monolingual adults also produced these structures, they cannot be claimed non-target like (however, see Baten and De Cuypere, 2014, for a different perspective on ungrammatical PP structures). For a bilingual's example, see Example 7a and for a monolingual's example, see 7b.

(7) a. *ich gebe das pferd an den schwein*[2]
i give the horse DO-ACC to the pig IO-ACC
"i give the horse to the pig"

b. *ich schenke die schlange auch an die schnecke*
i give the snake DO-ACC also to the snail IO-ACC
"i give the snake to the snail, too"

Moreover, utterances including verbs other than *geben/schenken* (to give) were excluded from the analysis since they may select different subject and object structures with different case assignments (see, e.g., *gehen an* [to go to] in Example 8).

(8) *der fisch geht an das schaf*
the fish SUBJ-NOM goes to the sheep DO-ACC
"the fish goes to the sheep"

Data were analyzed using generalized linear mixed models (GLMMs) with a logit link function, assuming a binomial error distribution for the binary response (correct/incorrect production). For the analysis, we used lme4 package (Bates et al., 2015) in R version 3.6.2 (R Core Team, 2020) and RStudio version 1.2.5033 (RStudio Team, 2020). Accuracy was analyzed split for the production of the accusative and dative objects. First, we set up a model space with all reasonable models, including a null model, just including *Participant* as a random intercept. Then, models were compared using the Akaike Information Criterion (AIC[3]). For the purpose of model comparisons, all models were computed using the Maximum Likelihood estimation method. The statistics of the winning model (i.e., best performing model according to AIC) were reported in full.

To evaluate whether we should include a factor *Group* dissociating

a) monolingual vs. bilingual participants only,
b) participants according to their age of German acquisition onset: from birth as L1 (monolinguals), from birth as 2L1, early L2 onset, late L2 onset,
c) participants based on the L1 case systems similar or different to German: same (monolinguals), similar, different,

we compared three GLMMs including different versions of *Group* as a fixed-effects factor, plus random-effects factor *Participant* as random intercept.

When categorizing the languages, we focused solely on marking of objects in the form of common nouns. Pronominal forms were not considered as only the use of full DPs (with definite articles) was relevant in this production task as well as in the following experiments (study 2 and study 3). We categorized a language as similar to the German case system when a case system with different cases used for subjects and objects as well as for the two objects of ditransitive structures was present. The categorization of case systems' similarity was specifically based on the presence of dative structures like in the German case system. All case systems of the languages categorized as similar, namely Albanian, Polish, Russian, Serbian and Turkish, have five to seven grammatical cases, among which are the accusative and the dative case. In Arabic, there are three grammatical cases. However, unlike in German, only the accusative is used for the object of a verb and it can mark both objects in double object constructions (Mohamed, 2013). Therefore, Arabic was classified dissimilar. English, Italian, Kurdish, Spanish, and Vietnamese were also categorized as dissimilar.

Production Results

To evaluate the question whether *Group* and *WordOrder* had an influence on production accuracy, we ran separate GLMMs for produced accusative and dative objects.

Accusative Production Accuracy

As described under 3.2, we first evaluated the different ways to model *Group*. The GLMM performing best was the one where bilinguals were further split according to AOO (AICs: 130.51 for mono-/bilingual split vs. 130.30 for age of acquisition onset split vs. 130.37 for case system split), but the differences between models were rather small. We continued with *Group* as split by AOO.

Second, we tested the influence of *Group* as defined above and *WordOrder* {IO-DO, DO-IO} on accuracy of the accusative object. To this end, we set up a model space as reported in **Table 3**. The winning model was the null model M0. The random intercept for *Participant* was highly significant [odds ratio = 25069.66, 95% CI = [507.19, 1239166.04], $p < 0.001$].

In other words, neither *Group*, that is, AOO, nor *WordOrder* helped explaining the variance in the data. The data were best explained by inter-individual variation only. This finding is not surprising given that monolinguals and bilinguals showed ceiling performance.

The few occurrences of errors in production within the bilingual group (see **Table 4**) were consistently accusative overgeneralizations in obligatory dative case contexts (see Example 9).

(9) *ich gebe die kuh den hund*
igive the cow ACC the dog ACC
"i give the cow (to) the dog"

Dative Production Accuracy

Again, we first evaluated how to best model the factor *Group*. The comparison of the monolingual/bilingual vs. AOO vs. case marking system of L1 models revealed that again AOO was the slightly better estimate (AICs = 95.79 vs. 95.64 vs. 95.77). Thus, we continued with Group as AOO.

[2]In this example, the bilingual speaker used the wrong gender for *Schwein* (pig), which is neuter in German instead of the produced masculine here.

[3]The AIC is an estimation of the relative quality of statistical models, considering a trade-off between the goodness of fit and simplicity of the model. Generally speaking: The model with the smallest value within a comparison fits the data best.

TABLE 3 | Model space for accuracy in the production task.

Model	Formula	AIC	
		Accusative	Dative
M0	Accuracy ~ 1 + (1|Participant)	**128.71**	93.80
M1	Accuracy ~ 1 + Group + (1|Participant)	130.30	95.64
M2	Accuracy ~ 1 + WordOrder + (1|Participant)	129.39	**93.63**
M3	Accuracy ~ 1 + Group + WordOrder + (1|Participant)	133.53	95.54
M4	Accuracy ~ 1 + Group + WordOrder + Group*WordOrder + (1|Participant)	-	97.36

Formulas for all models with AICs, i.e., estimators of model fit for accusative and dative case marking accuracy, respectively. Winning models' AICs are printed in bold letters. Model 4 could not be estimated for accusative due to an error in the Hesse matrix (1 negative eigenvalue), suggesting that there was too little variance in the data for a third term in the model—as to be expected from the vanishingly low error rate.

TABLE 4 | Performance of ditransitives in monolinguals and bilinguals in the production task.

	Accusative		Dative	
	IO-DO	DO-IO	IO-DO	DO-IO
Monolinguals				
Mean (%)	100	100	100	98.9
SD (%)	0.0	0.0	0.0	3.0
Range (%)	100–100	100–100	100–100	87.5–100
Bilinguals				
Mean (%)	100	97.3	100	99.6
SD (%)	0.0	8.7	0.0	1.6
Range (%)	100–100	65–100	100–100	93.7–100

Then, we tested the influence of *Group* as defined above and *WordOrder* {IO-DO, DO-IO} on accuracy of the dative object with the same models as for the accusative (see **Table 3**). Surprisingly, the model M2, containing a fixed-effects factor for *WordOrder* slightly outperformed the null model M0. The random intercept of *Participant* was highly significant [odds ratio = 7175.65, 95% CI = [93.74, 549265.32], $p < 0.001$]. However, even though slightly contributing to explaining the variance observed in the data, main effect of *WordOrder* did not reach significance [$\beta = 1.412$, SE =1.065, odds ratio = 4.10, 95% CI = [0.51, 33.03], $z = 1.326$, $p = 0.185$]. It is conceivable that this is due to the very low number error rates, thus, little variance in the data.

Thus, while *Group* had no influence on dative production accuracy, *WordOrder* had a descriptive, but non-significant influence.

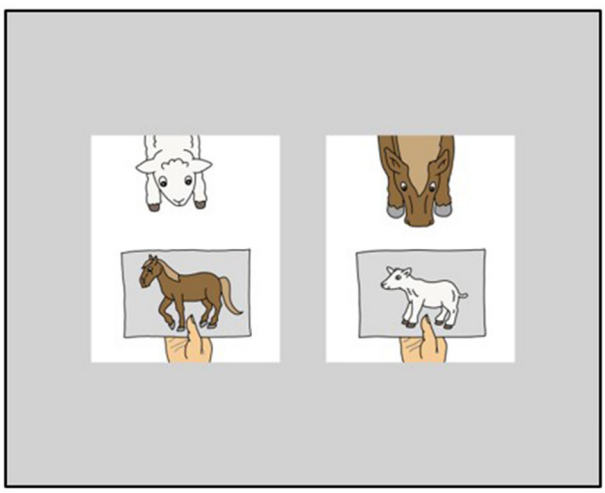

FIGURE 1 | Item example as presented during the sentence-picture matching task "ich gebe dem lamm sicherlich das pferd" ("certainly, i give the horse to the lamb").

STUDY 2 – SENTENCE-PICTURE MATCHING TASK (COMPREHENSION)

Comprehension Study Design

This tasks' stimuli were presented on a standard desktop computer running Windows 10. We used a 24" flatscreen monitor with a resolution of 1366×768 pixels and a frame rate of 60 Hz. The experimental presentation was conducted using a video created with Microsoft PowerPoint 2016.

Two pictures (408×491 pixels each) were simultaneously presented to the participants at a distance of 100 pixels in the middle of the screen on a gray background (see **Figure 1**). Participants heard a pre-recorded ditransitive construction via video (see Example 10), and were asked to press one of two

buttons as soon as they had identified the matching picture: the left (key "a") or right button (key "6" of the number block) corresponding to the side of the matching picture on screen. These buttons were chosen, because these keys are at a comfortable distance on a German QWERTZ keyboard with a number block. Both keys were marked with white stickers.

Regarding the stimuli of the comprehension task, we controlled for animacy, definiteness and verb bias as was described above for the production task. We only included objects that were [+animate], definite and consisted of full lexical DPs. To control for auditory length in addition, we restrained the maximal word length of objects to two syllables (most animals were monosyllabic). The subjects were kept consistent in 1st person singular for test trials. Regarding grammatical gender, we excluded the feminine gender from the comprehension task because of the homonym form of *der* for dative feminine singular and nominative masculine singular; since the latter is the default, the dative feminine *der* could be biased. We further excluded masculine items because of the difficulty to discriminate accusative and dative forms of *den* and *dem*. Therefore, all the objects included in the sentence-picture matching task were of neuter gender (*das schwein* [the pig], *das pferd* [the horse], *das schaf* [the sheep], *das pony* [the pony], *das lamm* [the lamb]). In addition, we added a task-irrelevant adverb between the indirect and direct object to provide time to parse the first object and react before hearing the second object. Example 10 illustrates the item composition. **Figure 1** presents examples of pictures, one of which had to be chosen for this item (target picture on the left, distractor on the right).

(10) *ich gebe dem lamm sicherlich das pferd*
i give (to) the lamb IO-DAT certainly the horse DO-ACC
"certainly, I give the horse to the lamb"

The experiment contained 58 items (20 ditransitive experimental trials and 38 fillers; e.g., *die giraffe frisst auf der wiese* [the giraffe eats in the meadow] vs. *das schaf frisst auf der wiese* [the sheep eats in the meadow], see item list in **Supplementary Table 2**). Prior to the experiment, four practice items were given. The experiment started only after the participants understood the task. We created two lists of items. Each list contained all experimental trials and all fillers but their order was different. The two different lists of pseudorandomized item order were assigned to the participants. Therefore, items were numbered and then randomized using a random number generator (www.random.org). After that, we manually adapted the resulting random order because of task-related issues. When a test trial followed immediately after a break, we changed the order and put a filler item instead (see **Supplementary Table 2** for the item composition). In total, this experiment lasted around 10 min.

Comprehension Data Analysis

We analyzed accuracy of sentence-picture matches and RTs. For RTs, the annotation capture plugin of the software Pupil Labs was used. This plugin allows for labeling timestamps ("L" for left picture, "R" for right picture). These labels are created by pressing their respective hotkey (as for the sentence-picture matching task, we chose "a" for left and "6" for right). Regarding the RT analysis, we defined critical windows to investigate how fast participants would react after hearing the first case marking (see **Figure 2**). RTs were then calculated by subtracting the timestamp of the critical window's starting point from the timestamp of the label L/R. As not all of the items had precisely the same length, critical windows were defined individually for each of the 20 trials by using the software Audacity® version 2.2.2 (iWeb Media, Ltd., Birkirka, Malta). Accuracy was analyzed by comparing the participants' response (L/R) and the correct target response.

Regarding predictions, we followed Schlenter (2019, p. 2) in her assumption that "only effects visible prior to the onset of the critical perceptual input are taken as effects of prediction" in contrast to later effects that may reflect rapid integration rather than prediction. As this study instrumentalizes RTs as indicators of anticipation, we defined the time between the offset of the object's article and the onset of the second object as the critical window. Trials in which participants reacted before the onset of the second object in this critical window were classified as "predictions." Trials in which participants reacted after the offset of the critical window were classified as "no predictions." Since not all trials have precisely the same length, the reaction time at the participants' disposal before hearing the second object varies between trials ($min = 1{,}471$ ms, $max = 1{,}999$ ms). Therefore, prediction ratings were scored individually for each item.

Data were analyzed using linear mixed-effects models (LMMs), assuming a Gaussian distribution for interval-scaled RTs and GLMMs with a logit link function for binary accuracy with lme4 package (Bates et al., 2015) in R version 3.6.2 (R Core Team, 2020) and RStudio version 1.2.5033 (RStudio Team, 2020). For the analysis of RTs, erroneous trials were excluded. Then, we set up a model space with all reasonable models, including a null model, just including *Participant* as a random intercept. Models were compared using the AIC. For the purpose of model comparisons, Maximum Likelihood estimation method was used. The statistics of the winning model (= lowest AIC) are reported in full in the text.

Comprehension Results

Comprehension Accuracy

The results of the participants' accuracy scores in the sentence-picture matching task are listed in **Table 5**. As can be seen, with means >95% in every group and both word orders, accuracy scores are consistently high.

To evaluate whether we should include a factor *Group* dissociating

a) monolingual vs. bilingual participants only,
b) participants according to their age of German onset: from birth as L1 (monolinguals), from birth as 2L1, early L2 onset, late L2 onset,
c) participants based on the L1 case systems similar or different to German: same (monolinguals), similar, different,

we again first compared three GLMMs including different versions of *Group* as a fixed-effects factor, plus random-effects

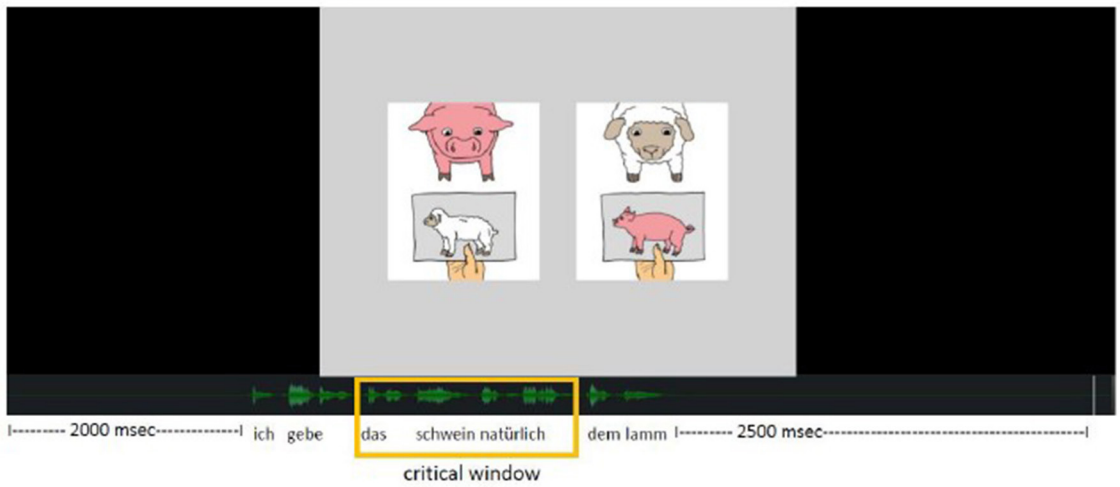

FIGURE 2 | Critical Window for analysis of the reaction times in the sentence-picture matching task. Item example "ich gebe das schwein natürlich dem lamm" ("of course, i give the pig to the lamb").

TABLE 5 | Monolinguals' and bilinguals' accuracy in the sentence–picture matching task.

	IO-DO	DO-IO	Total
Monolinguals			
Mean (%)	97.3	97.3	97.3
SD (%)	5.9	5.2	3.7
Range (%)	80–100	80–100	90–100
Bilinguals			
Mean (%)	96.7	95.6	96.1
SD (%)	8.2	10.1	7.6
Range (%)	70–100	60–100	75–100
Mann-Whitney-U-test	$p > 0.05$	$p > 0.05$	$p > 0.05$

factor *Participant* as random intercept. The results were clear: The differences were negligible (AICs: 1534.87 vs. 1534.88 vs. 1534.90). We decided to continue with version (a) of factor *Group*, because of its lowest AIC.

Then, to evaluate the different influences of *Group* {monolingual, bilingual} and *WordOrder* {IO-DO, DO-IO}, we set up different models as described in **Table 6** and compared them using the AIC. As to be expected from the ceiling performance of both mono- and bilinguals (i.e., above 90%; see **Table 5**), no model performed better than the null model M0 (see **Table 6**).

Thus, regarding the accuracy in the sentence-picture matching task, we did not find any differences aside from the ones that were due to inter-individual differences based on random-effects factor *Participant* [odds ratio of intercept = 68.52, 95% CI = [28.03, 167.53], z = 9.27, $p < 0.001$].

Reaction Times

Reaction times were similar in monolinguals and bilinguals (see **Figure 3**). For the analysis, we again compared three LMMs including different versions of *Group* as a fixed-effects factor, plus random-effects factor *Participant* as random intercept. The comparison of the monolingual/bilingual vs. AOO vs. case marking system of L1 models revealed that AOO was the slightly better estimate (AICs: 1437.41 vs. 1437.37 vs. 1437.40). Again, the differences in AICs are marginal. Nevertheless, we continued by modeling *Group* as AOO.

Then, to evaluate the different influences of *Group* {monolingual, bilingual} and *WordOrder* {IO-DO, DO-IO}, we set up different models as described in **Table 6** and compared them using the AIC. As can be seen from **Table 6**, model M2 was the winning model. Random intercept for *Participant* was significant [$\beta = 1.728$, SE = 0.086, 95% CI = [1.56, 1.90], $t(47.99) = 20.117$, $p < 0.001$] and fixed-effects factor *WordOrder* was highly significant [$\beta = -0.098$, SE = 0.035, 95% CI = [−0.17, −0.03], $t(808.19) = -2.78$, $p = 0.006$].

Mean RT for IO-DO (1.637 s, SD = 0.726 s) was smaller than for DO-IO (1.723 s, SD = 0.773 s), displaying a typical word-order effect.

Anticipation of the Second Object

Regarding anticipatory reactions, we evaluated the influence of *Group* and *WordOrder* on whether participants were able to anticipate the upcoming second object. Only correct

TABLE 6 | Model space for accuracy and RTs in the comprehension task.

Model	Formula	AIC	
		y = accuracy	y = RT
M0	y ~ 1 + (1|Participant)	**232.33**	1435.43
M1	y ~ 1 + Group + (1|Participant)	234.22	1437.37
M2	y ~ 1 + WordOrder + (1|Participant)	233.32	**1429.75**
M3	y ~ 1 + Group + WordOrder + (1|Participant)	235.20	1431.68
M4	y ~ 1 + Group + WordOrder + Group*WordOrder + (1|Participant)	237.16	1432.98

Formulas for all models with AICs, i.e., estimators of model fit for the sentence-picture matching task. The winning models' AICs are printed in bold font.

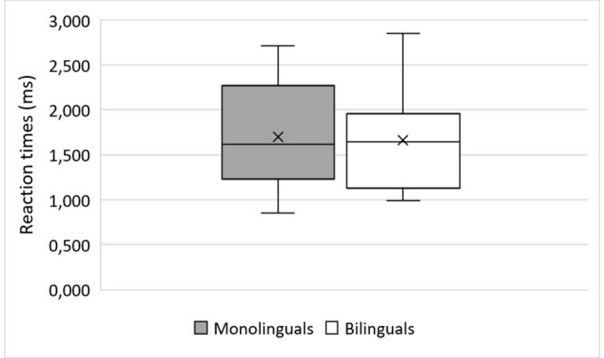

FIGURE 3 | Reaction times after hearing the onset of the first object in the sentence-picture matching task in monolingual and bilingual speakers. x = mean (mean monolinguals = 1,700 ms, SD = 0,561 ms; mean bilinguals = 1,663 ms, SD = 0,562 ms).

responses were analyzed. First, we assessed how to best model *Group* by comparing the three different assessments as before. Again, all model fit estimates (AICs) were relatively similar (monolingual/bilingual = 830.12, AOO = 830.21, case similarity = 830.23). As the simple differentiation only based on mono- vs. bilingual was the best fit, we used this for Group.

We then assessed which GLMM fit the data best with regard to whether *WordOrder* and *Group* had explanatory value for anticipation. The full model space is provided in **Table 7**. The model that fit the data best was the null model M0. However, even random intercept for *Participant* was not significant [odds ratio = 1.44, 95% CI = [0.72, 2.90], $p = 0.302$].

Thus, neither *Group* nor *WordOrder* appear to influence the number of anticipated second objects.

In accordance with this, **Figure 4** shows that no significant differences were found between the mean percentage of anticipations after hearing the first object in monolinguals (mean = 55.0%, SD = 34.1%) and bilinguals (mean = 60.0%, SD = 34.2%; Mann-Whitney U: U = 559.5, p =0.541). Thus, bilinguals predicted the second object to the same extent, that is, in more than 60% of all items, as monolinguals.

Lastly, there were no significant correlations between any of the tested WM categories and overall rate of anticipatory reactions (see **Table 8**).

STUDY 3 – PUPILLOMETRY

Pupillometry Study Design

Pupil data were tracked by a Pupil Labs eye tracker at 200 Hz (*Pupil Core*, Pupil Labs). We used a 9-point eye tracker calibration before the pupillometry task started. To avoid confounding implicit pupil measurements with a participant's explicit action, i.e., manual response, the participant's task was simply to listen to the auditory stimulus. To minimize reflexive reactions of the pupil diameter to changes in luminance, we kept the displayed colors constant. Therefore, the monitor displayed a black fixation cross on a gray background while auditory stimuli were played. After a block of five trials, there was either a comprehension question to check for the participants' attention (like e.g., "Is today Wednesday?") or pause with a picture of a forest to relax the eyes. The only explicit reaction that was required by the participants throughout the experiment was to press a button (yes/no) to answer these comprehension questions. Throughout the presentation of the forest picture, the participants were told to do whatever they like to relax the eyes (blink, look at the picture, or look away). The stimuli following the comprehension questions and the stimuli following the relax pictures were fillers that were not analyzed.

The experiment contained 30 test items and 20 filler items such as *das schaf frisst das gras* ("the sheep eats the grass," see **Supplementary Table 3** for further examples). The test items were constructed similarly to the items from the comprehension task and contained 10 items per condition (A, B, C, see Example 11) resulting in a total amount of 30 grammatical stimuli (condition A + fillers) and 20 ungrammatical stimuli (conditions B and C). Conditions contained gradually violated grammars: Whereas condition A contained grammatical items (see Example 11a), condition B contained an accusative overgeneralization in the dative context (see Example 11b). This equals the prefinal acquisition step toward target ditransitive production in child acquisition. Condition C contained no determiners at all (see Example 11c), which equals the first step in the acquisition process. Condition B therefore was labeled "slightly ungrammatical" and condition C was labeled "strongly ungrammatical." In total, the pupillometry experiment lasted about 9 min.

(11) a. *Condition A*
ich gebe die kuh sicherlich der giraffe

TABLE 7 | Model space for anticipations of the second object in the comprehension task.

Model	Formula	AIC
M0	Anticipation ~ 1 + (1|Participant)	**828.24**
M1	Anticipation ~ 1 + Group + (1|Participant)	830.12
M2	Anticipation ~ 1 + WordOrder + (1|Participant)	829.71
M3	Anticipation ~ 1 + Group + WordOrder + (1|Participant)	831.59
M4	Anticipation ~ 1 + Group + WordOrder + Group*WordOrder + (1|Participant)	833.02

The winning model's AIC is printed in bold font.

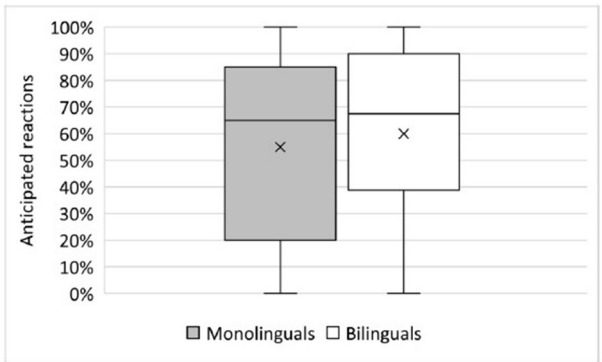

FIGURE 4 | Proportion of anticipatory reactions (before hearing the second object) in the sentence-picture matching task in monolinguals and bilinguals. x = mean.

i give the cow DO-ACC certainly (to) the horse IO-DAT
"certainly, i give the cow to the giraffe"
b. *Condition B*
**ich gebe das lamm sicherlich das schaf*
i give the lamb DO-ACC certainly the sheep IO-ACC
"certainly, i give the lamb the giraffe"
c. *Condition C*
**ich gebe pferd sicherlich schwein*
i give (the) horse certainly (the) pig
"certainly, i give lamb giraffe"

Pupillometry Data Analysis

Pre-processing was accomplished using the package gazeR (Geller et al., 2020) in R (version 1.4.1106, R Core Team, 2020). Samples 100 ms prior and 100 ms after a blink were coded as missing and were linearly interpolated. After the interpolation process, artifacts were removed based on the median absolute deviation (see Geller et al., 2020 for details). Such artifacts stem from quick changes in pupil size. In order to smooth the pupil time course, we passed a 5-point moving average over the data. In order to account for spontaneous variation in pupil size, we baseline-corrected the pupil diameter for each trial individually. Therefore, we determined a baseline of 1,000 ms prior to audio onset. Mean pupil diameter from this baseline was then subtracted from all pupil data points of the respective trial. A time window of 3,500 ms starting from the first violation (i.e., the onset of the first object) was selected for analysis. Moreover, data were filtered for intra-individual outliers by filtering all data points with a pupil diameter +/− 2.5 SD above/below the mean.

In the absence of a "field-standard statistical approach" (Geller et al., 2020, p. 2251) to analyze pupil data, we decided to analyze the pupil dilation trajectories directly instead of extracting peak amplitudes and latencies, as recommended by van Rij et al. (2019). We therefore applied a non-linear regression analysis, i.e., generalized additive mixed modeling (GAMM). To this end, we used mgcv package version 1.8.36 (Wood, 2017; van Rij et al., 2019) in R version 4.0.4 (R Core Team, 2020) and RStudio version 1.4.1106 (RStudio Team, 2020).

Pupillometry Results

To find the best model fitting the data, we set up a model space including different combinations of fixed-effects factors and random smooths as elaborated below. The full model space is provided in **Table 9**, where *y* is the baseline-corrected, outlier-corrected (+/− 2.5 SD of the mean) pupil dilation. All models were estimated using Maximum Likelihood estimation.

Furthermore, all models include a covariate smooth term for gaze direction s (x,y), because the measured pupil dilation is confounded with it. Depending on the angle, the pupil appears oval instead of round, leading to an underestimation of the pupil size. The null model M00 further includes a random smooth for *Participant* over *Time s* {time, participant} to model the variance of the pupil response over time that is only due to inter-individual differences. We chose to also estimate a second null model M0, which also included a random smooth *s* {time, trial}to account for trial-wise variation over time. Because M0 was the better null model, we included both random smooths in all models of interest (M1–M3, **Table 9**).

Models M1–M3 further included either a fixed-effects factor to model main effect of *Violation* {A, B, C}, *Group* {monolingual, bilingual}, or both *VioGr* {A-monolingual, A-bilingual, B-monolingual, B-bilingual, C-monolingual, C-bilingual}. Due to the nature of GAMMs, an interaction term cannot be included as one would do for LMMs. Thus, we included a factor *VioGr* combining all levels of *Violation* and *Group*. The best fitting model by far was M3, the one including the factor *VioGr*.

TABLE 8 | Results of correlation analyses between working memory (forward [FW] and backward [BW] digit span) and predictive processing (in percentage of anticipatory reactions).

	Monolinguals	Bilinguals
Digit span FW–Anticipatory reactions (%)	$r = 0.02$	$r = 0.44$
	$p > 0.05$	$p > 0.05$
Digit span BW–Anticipatory reactions (%)	$r = -0.14$	$r = 0.06$
	$p > 0.05$	$p > 0.05$

TABLE 9 | Model space for pupil dilation in the pupillometry task.

Model	Simplified R formula	AIC
M00	y ~ 1 + s(x, y) + s(time, participant)	272547.51
M0	y ~ 1 + s(x, y) + s(time, subject) + s(time, trial)	268361.61
M1	y ~ 1 + violation + s(x, y) + s(time, by = violation) + s(time, subject) + s(time, trial)	148714.02
M2	y ~ 1 + group + s(x, y) + s(time, by = group) + s(time, subject) + s(time, trial)	149523.24
M3	y ~ 1 + viogr + s(x, y) + s(time, by = viogr) + s(time, subject) + s(time, trial)	**147441.16**

The winning model's AIC is printed in bold font.

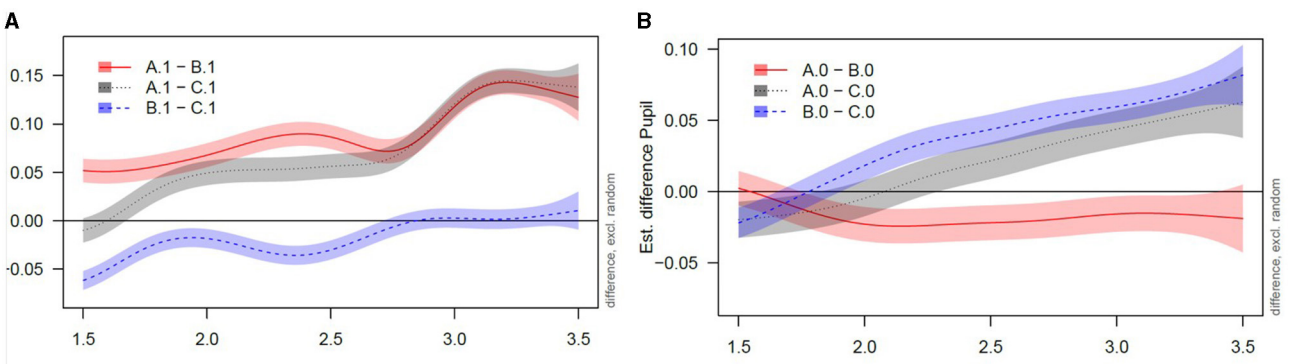

FIGURE 5 | Monolinguals' and bilinguals' pupil dilation in the pupillometry task. Pupil dilation within the analysis window of 3,500 ms (time zero starting at the onset of the first object) with plotted differences between conditions. Curves around 0.00 indicate no difference between the plotted conditions (e.g., between A0 and B0). Curves deviating from 0.00 indicate differences between the plotted conditions in the pupil dilation trajectory. A1 = condition A (grammatical) in monolingual data; A0 = condition A (grammatical) in bilingual data; B1 = condition B (slightly ungrammatical) in monolingual data; B0 = condition B (slightly ungrammatical) in bilingual data; C1 = condition C (strongly ungrammatical) in monolingual data; C0 = condition C (strongly ungrammatical) in bilingual data. Panel **(A)**: Monolinguals. Panel **(B)**: Bilinguals.

In this task the pupil dilation is expected to change with violation (condition B and C vs. condition A). As to be expected from the winning model M3, excluding random effects of *Participant* and *Trial*, pupil dilation varied as a function of *Violation* {A, B, C} and *Group* {monolingual, bilingual}. This is depicted in **Figure 5**.

As **Figure 5** illustrates, both groups react differently to different conditions. As suggested in van Rij et al. (2019), we inspected the model's estimates of the differences between the conditions visually and therefore decided to interpret the time window where curves of differences obviously change (from around 1.50 s after the time zero for both groups, see **Figure 5**). To be temporally exact, for monolinguals, the difference of the pupil dilation for grammatical sentences (A) and slight violations of type B was significant between 1.50 and 3.50 s. Similarly, pupil dilation in response to grammatical sentences (A) and violations of type C showed significant differences between 1.77 and 3.50 s. Both violation conditions (B vs. C) on the other hand differed significantly between 1.50 and 2.83 s and then converged. Contrarywise, bilinguals showed significant differences between grammatical sentences (A) and violations of type B between 1.67 and 3.50 s, and for A and C, a difference was observed only later between 2.68 s and 3.50 s. In bilinguals, the two violation conditions B and C differed between 2.02 and 3.50 s. Thus, B and C did not converge as in monolinguals. Therefore, the main difference between monolinguals and bilinguals is the bilinguals' relatively late reaction on strongly ungrammatical sentences (2.68 s in bilinguals vs. 1.77 s in monolinguals). The differences in pupil dilation time courses (as a covert or implicit response) for mono- and bilinguals are especially interesting in light of the missing behavioral differences, that is, of the overt response. This will be discussed below.

DISCUSSION

This paper aimed to investigate monolingual and bilingual adult speakers of German with respect to their linguistic performance regarding production, comprehension, and implicit sensitivity to grammatical violations of the morphosyntactic complex double-object construction with ditransitive verbs. Therefore, we developed three experiments in which we elicited ditransitive structures (study 1), assessed accuracy and RTs in a sentence-picture matching task (study 2), and investigated sensitivity to grammatical violations via pupillometry (study 3).

In comprehension and production, bilinguals exhibited abilities that were comparable to monolinguals' in all aspects. Regardless of the AOO (0;0 for 2L1, 3;0–4;0 for eL2 and 6;0–13;0 for late L2), all bilinguals performed at ceiling in production and comprehension tasks. The reported overgeneralisations of accusative in dative contexts in production were negligible and can be interpreted as performance errors. However, since *WordOrder* had a descriptive, but non-significant influence on the production of the dative it would be interesting to see, whether the effect might reach significance, when performance is not at ceiling.

The ability to anticipate upcoming input, i.e., using the case marking of the first object (dative marking in default word order and accusative marking in marked word order) to anticipate the upcoming input in bilinguals was comparable to that in monolingual speakers. This is in accordance with Schlenter's (2019) findings and contradicts the RAGE hypothesis, which states that bilinguals have a reduced ability to predict upcoming input (Grüter et al., 2014). The variation in RTs within the bilingual group could not be explained by different AOOs. Thus, we did not observe that 2L1 speakers reacted faster than monolinguals (as it was the case in Desideri and Bonifacci, 2018). However, a bigger sample size is needed to generalize these preliminary results. A more trivial explanation for the lacking difference in our findings between monolinguals and bilinguals may be that the task was too easy for the adult participants to reveal subtle processing differences, owing to ceiling effects. This potential limitation could be addressed in follow-up studies that employ tasks that are more difficult. However, the aim of such a study would stand to question. There are probably no real-world implications if a task has to be extremely difficult for differences to emerge between bilinguals and monolinguals.

As can be seen in the huge variability in monolingual and bilingual adults regarding anticipated reactions, the methodological set-up of this study limits definite conclusions about the underlying predictive ability. Nevertheless, the data show comparable levels of anticipated reactions for mono- and bilinguals, supporting the overall impression of bilinguals' native speaker competence. This is in line with Halliday's (1975) definition of a native speaker as someone who is able to predict what the other person is going to say and therefore being able to anticipate upcoming input. Furthermore, we found a word order effect in the comprehension task that concerned the strategy of "IO-first." There was a word order bias in favor of the IO-DO order in the sense that speakers reacted significantly faster in trials with unmarked IO-DO word order than in the marked DO-IO word order. This finding is in line with Kholodova and Allen (in press) report on the productive preference of IO-DO word order in adult native German speakers. In accordance with the N1 bias found for subjects, where children up to puberty implicitly prefer interpreting the first NP in an utterance as the subject (Lidzba et al., 2013), this can be interpreted as a processing strategy. The investigated speakers implicitly assume that the first NP is the subject and the second NP is the recipient (i.e., the IO in ditransitives), which in most naturally occurring cases leads to the correct utterance interpretation. Applying this strategy may enhance utterance interpretation. Both, monolingual and bilingual speakers show this robust word order effect in the sentence-picture matching task.

The analysis of the participants' pupil data revealed a clear difference between the implicit response to grammatical (condition A) compared to slightly ungrammatical (condition B) and strongly ungrammatical sentences (condition C) in both, monolinguals and bilinguals. Taking, for instance, Chomsky's relation between native speaker and a grammatical sentence as a basis, the participants investigated here can be claimed highly competent since they were able to identify ungrammatical utterances implicitly and thus provide valid implicit grammaticality judgements on their language on the basis of intuitive knowledge of the grammatical sentence (Chomsky, 2002).

However, statistical analyses revealed that pupil dilation varied not only as a function of *Violation* {A, B, C} but also as a function of *Group* {monolingual, bilingual}. The main difference between monolinguals and bilinguals in the timing of pupil reactions concerns the difference between A and C, i.e., the grammatical condition and the strongly ungrammatical condition. Here, monolinguals react faster to the grammatical violation than bilinguals, demonstrated by the differences between A and C that become significant earlier in monolinguals than in bilinguals. The bilinguals' delayed response when compared to monolinguals could be due to higher processing efforts. This interpretation would be in line with Fernandez (2016) who interprets longer peak latencies (i.e., the time until the pupil is maximally dilated) as measures for higher processing effort.

Overall, despite bilinguals' comparable production and comprehension ability evident from their behavioral responses, the pupil dilation data revealed a temporal difference in violation detection for strong grammatical errors. This in turn indicates a subtle difference between monolingual and bilingual speakers' reactions to these strong grammatical violations. This could be interpreted as confirming Paradis' (2016, 2019) findings of bilingual speakers being typically indistinguishable from monolingual speakers in conversation, while they differ when it comes to grammaticality judgement tasks regarding complex morphosyntax in adulthood. However, our data revealed bilinguals to be able to identify these violations but with a different pace than monolinguals. Therefore, we argue for an intact although somewhat delayed implicit identification ability of grammatical violations and assume that there is no real-life effect of this delay in the millisecond range.

However, a further methodological note is important at this point. Even though the AIC differences were rather small, the

AICs for three of the models we calculated for production and comprehension data indicated that some of the variance is better explained by a Group variable that does not only differentiate between monolinguals and bilinguals, but that differentiates bilinguals further based on their AOO. Thus, with a larger sample size and more data from highly proficient bilinguals with different AOOs, we might be able to detect a potential AOO effect that is concealed in our data by the small[4] sample sizes of bilingual subgroups with different AOOs. This is up to future research.

Concluding Remarks

Overall, the results of this pilot study demonstrate high proficiency in production as well as comprehension in the studied group of bilingual adult speakers of German regardless of their (2)L1 and AOO, evident in their comparable ceiling performance in production and comprehension, as well as their high sensitivity to grammatical violations in German utterances. We therefore conclude that at this proficiency level, AOO and cross-linguistic influence may not affect production and comprehension abilities in bilingual adult speakers of German. The documented high proficiency in all domains is in line with Hartshorne et al. (2018), who proposed the sensitive period up to age 17. We did not find support for a turning point of grammar learning abilities at a younger age (contrasting Meisel, 2018). Regarding ultimate attainment, this indicates that even after a relatively late AOO (13 years), complex morphosyntactic structures such as ditransitives can be mastered in production and comprehension to a monolingual and 2L1 native-speaker degree. Conversely, bilinguals have the advantage of having acquired an additional language while showing the same explicit and implicit grammatical knowledge of complex structures like monolinguals. However, a possible limitation is the small sample size of the current study. The results do not imply that every eL2 child can master case marking in ditransitives, but they at least provide evidence that early and even late L2 learners of German can master these rather complex structures in morphosyntax, and thus reach a competence level comparable to monolinguals or 2L1 speakers.

In line with psycholinguistic research on early bilingualism, we included a control group of monolingual speakers in the present study. Our aim was to challenge the idealized monolingual L1 competence as the native speaker norm and evaluate the validity of this assumption. Our findings provide evidence for a new perspective on the term "native speaker." Here, our data revealed that even so-called near-native speakers with AOOs later than 6 years could show high proficiency throughout. Characteristics previously attributed to native speakers (i.e., prediction of upcoming input in Halliday, 1975, and valid grammaticality judgements by Chomsky, 2002) were also found to be true for the bilingual speakers investigated in the present study. Therefore, the connotation of the term "native" with high proficiency and "non-/near-native" with lower proficiency appears to be misleading. The term "near-native" suggests that someone who is not born in the country of the L2 is only near to the competence of a native speaker (see Bylund et al., 2021), whereas "native" is equated with monolingualism, and monolingualism in turn is equated with the highest proficiency. However, the huge variation in the RTs and in the pupillometry measures of the investigated monolinguals of our study convey a different picture: Even within highly proficient monolingual native speakers, there is considerable variance with regard to language skills. If the explicit and implicit competence of bilingual speakers (who at random have not been raised with the language under investigation from birth) fall within this spectrum of high proficiency, then it is not reasonable to call them near-native, but instead to focus on their competences, which may very well be at monolingual native level. The term "near-native" is connoted with imperfection that could not be documented in our investigated bilingual speakers.

The term "native speaker" is misleading in the sense that being raised monolingually does not mean being perfect in every subtle part of one's own language (see high in-group variance in RTs). On the other hand, acquiring an additional language or being raised bilingually does not mean that someone is not able to achieve high competence, also in subtle and implicit measures. We therefore call for a different perspective on someone's language competence other than to tie it invariably to the place of birth and upbringing and the amount of languages that have been acquired.

To this end, the two "groups" (monolinguals and bilinguals) that we compared and contrasted in our study can be collapsed in to one group: In fact, we see one group of highly proficient speakers of German with high performances on different explicit and implicit tasks that concern complex morphosyntax. With the add-on of some speakers who are additionally able to speak another language, representing a personal advantage for these speakers.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

A-LS developed the research ideas, the design of the study, and wrote the first draft of the manuscript. A-LS collected the data, with the help of JK. A-LS, JK, and TL performed the statistical analysis and created the visualizations. GU and JK wrote sections

[4]It should be noted that the study was not designed to further split bilinguals based on AOO. This modeling was an exploratory addition.

of the manuscript. All authors contributed to manuscript revision and approved the submitted version.

ACKNOWLEDGMENTS

We acknowledge financial support by Deutsche Forschungsgemeinschaft and TU Dortmund University within the funding programme Open Access Publishing. We acknowledge financial support by the University of Hildesheim for a student assistant. We thank Tom Fritzsche for helpful comments on the pupillometry experiment setup and Prof. Dr. Kristian Folta-Schoofs for allowing the usage of his lab and equipment. For their help with data collection, we thank Klara Tittelbach-Helmrich and David Mietzner. We further thank Hannes Elfers and the Elisabethschule in Hildesheim for their invaluable help with data collection and Julia Diedrich for illustrating the visual stimuli. Above all, we thank our participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717379/full#supplementary-material

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

Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1

Baten, K., and De Cuypere, L. (2014). The dative alternation in L2 German? conceptualization transfer from L1 Dutch. *Vigo Int. J. Appl. Linguist.* 11, 9–40. Available online at: https://revistas.webs.uvigo.es/index.php/vial/article/view/60

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 276–292. doi: 10.1037/0033-2909.91.2.276

Beatty, J., and Lucero-Wagoner, B. (2000). "The pupillary system," in *Handbook of Psychophysiology*, eds J. Cacioppo, L. Tassinary, and G. Berntson (Cambridge: Cambridge University Press), 142–162.

Bradley, M. M., Miccoli, L., Escrig, M. A., and Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 602–607. doi: 10.1111/j.1469-8986.2008.00654.x

Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. *Nature* 216, 515–516. doi: 10.1038/216515a0

Brandt, S., Lieven, E., and Tomasello, M. (2016). German children's use of word order and case marking to interpret simple and complex sentences: testing differences between constructions and lexical items. *Langu. Learn. Dev.* 12, 156–182. doi: 10.1080/15475441.2015.1052448

Bylund, E., Hyltenstam, K., and Abrahamsson, N. (2021). Age of acquisition—not bilingualism—is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism Lang. Cogn.* 24, 18–30. doi: 10.1017/S1366728920000188

Chomsky, N. (2002). *Syntactic Structures, 2nd Edn*. Berlin: Mouton de Gruyter.

Clahsen, H., and Felser, C. (2006). How native-like is non-native language processing? *Trends Cogn. Sci.* 10, 564–570. doi: 10.1016/j.tics.2006.10.002

Cook, V. (2008). "Multi-competence: black hole or wormhole for second language acquisition research?" in Second *Language Acquisition: Vol. 25. Understanding Second Language Process*, eds Z. Han and E. S. Park (Bristol: Multilingual Matters), 16–26.

Davies, A. (2004). "The native speaker in applied linguistics," in *Blackwell Handbooks in Linguistics: Vol. 17. The Handbook of Applied Linguistics*, ed A. Davies (Oxford: Blackwell), 431–450.

Davson, H. (1972). *The Physiology of the eye*. Churchill Livingstone.

Desideri, L., and Bonifacci, P. (2018). Verbal and nonverbal anticipatory mechanisms in bilinguals. *J. Psycholinguist. Res.* 47, 719–739. doi: 10.1007/s10936-017-9556-1

Dewaele, J.-M. (2017). Why the dichotomy "L1 versus Lx user" is better than "native versus non-native speaker." *Appl. Linguist.* 39, 236–240. doi: 10.1093/applin/amw055

Dittmar, M., Abbot-Smith, K., Lieven, E., and Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Dev.* 79, 1152–1167. doi: 10.1111/j.1467-8624.2008.01181.x

Drenhaus, H. (2004). *Minimalism, features and parallel grammars: On the acquisition of german ditransitive structures* [doctoral dissertation]. University of Potsdam, Potsdam, Germany.

Eisenbeiss, S. (2002). *Merkmalsgesteuerter Grammatikerwerb: Eine Untersuchung zum Erwerb der Struktur und Flexion von Nominalphrasen [feature-driven structure building. A study on the acquisition of noun-phrase structure and inflection]* [doctoral dissertation]. Heinrich-Heine-Universität Düsseldorf, Dusseldorf, Germany.

Eisenbeiss, S., Bartke, S., and Clahsen, H. (2006). 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

Engelhardt, P. E., Ferreira, F., and Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Q. J. Exp. Psychol.* 63, 639–645. doi: 10.1080/17470210903469864

Fernandez, L. (2016). *Investigating word order processing using pupillometry and event-related potentials* [doctoral dissertation]. University of Potsdam, Potsdam, Germany.

Fernandez, L., Höhle, B., Brock, J., and Nickels, L. (2018). Investigating auditory processing of syntactic gaps with L2 speakers using pupillometry. Second *Lang. Res.* 34, 201–227. doi: 10.1177/0267658317722386

Friesen, D. C., Luo, L., Luk, G., and Bialystok, E. (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Lang. Cogn. Neurosci.* 30, 238–250. doi: 10.1080/23273798.2014.918630

Gamper, J. (2019). The role of case and animacy in bi-and monolingual children's sentence interpretation in german: a developmental perspective. *Open Linguist.* 5, 1–24. doi: 10.1515/opli-2019-0001

Geller, J., Winn, M. B., Mahr, T., and Mirman, D. (2020). Gazer: a package for processing gaze position and pupil size data. *Behav. Res. Methods* 52, 2232–2255. doi: 10.3758/s13428-020-01374-8

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Appl. Linguist.* 35, 483–503. doi: 10.1093/applin/amu013

Grimm, H., and Doil, H. (2006). *ELFRA: Elternfragebögen für die Früherkennung von Risikokindern, 2nd Edn*. Göttingen: Hogrefe.

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

Grüter, T., Rohde, H., and Schafer, A. J. (2014). "The role of discourse-level expectations in non-native speakers' referential choices," in *Proceedings of the 38th Annual Boston University Conference on Language Development*, eds W. Orman and M. J. Valleau (Boston, MA: Cascadilla Press), 179–191.

Grüter, T., Rohde, H., and Schafer, A. J. (2017). Coreference and discourse coherence in L2. *Linguist. Approaches Biling.* 7, 199–229. doi: 10.1075/lab.15011.gru

Halliday, M. (1975). "Learning how to mean," in *Foundations of Language Development: A Multidisciplanary Approach*, eds E. Lenneberg and E. H. Lenneberg (Paris: Academic Press and Unesco Press), 239–265.

Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (2018). A critical period for second language acquisition: evidence from 2/3 million english speakers. *Cognition* 177, 263–277. doi: 10.1016/j.cognition.2018.04.007

Hepach, R., and Westermann, G. (2016). Pupillometry in infancy research. *J. Cogn. Dev.* 17, 359–377. doi: 10.1080/15248372.2015.1135801

Hess, E. H., and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 1190–1192. doi: 10.1126/science.143.3611.1190

Hopp, H. (2015). Individual differences in the second language processing of object–subject ambiguities. *Appl. Psycholinguist.* 36, 129–173. doi: 10.1017/S0142716413000180

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Res.* 1626, 118–135. doi: 10.1016/j.brainres.2015.02.014

Johnson, E., Miller Singley, A., Peckham, A., Johnson, S., and Bunge, S. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Front. Psychol.* 5:218. doi: 10.3389/fpsyg.2014.00218

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

Just, M. A., and Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Can. J. Exp. Psychol.* 47, 310–339. doi: 10.1037/h0078820

Kaan, E. (2014). Predictive sentence processing in L2 and L1: what is different? *Linguist. Approaches Biling.* 4, 257–282. doi: 10.1075/lab.4.2.05kaa

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Kamide, Y., Scheepers, C., and Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from german and english. *J. Psycholinguist. Res.* 32, 37–55. doi: 10.1023/A:1021933015362

Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *Int. J. Psychophysiol.* 52, 7–21. doi: 10.1016/j.ijpsycho.2003.12.002

Kauschke, C., and Siegmüller, J. (2010). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS) [Patholinguistic Assessment of Developmental Language Disorders in German], 2nd Edn*. Munich: Elsevier.

Kholodova, A., and Allen, S. (in press). "The ditransitive alternation in german: structural preferences and verb bias effects," in *Ditransitive Constructions in Germanic Languages: Diachronic and Synchronic Aspects*, eds M. Röthlisberger, E. Zehentner, and T. Colleman (Amsterdam Benjamins).

Kosch, T., Hassib, M., Buschek, D., and Schmidt, A. (2018). "Look into my eyes: using pupil dilation to estimate mental workload for task complexity adaptation," in *CHI '18: CHI Conference on Human Factors in Computing Systems*, eds R. Mandryk and M. Hancock (Montreal, QC).

Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspect. Psychol. Sci.* 7, 18–27. doi: 10.1177/1745691611427305

Lawson, R. P., Mathys, C., and Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* 20, 1293–1299. doi: 10.1038/nn.4615

Lemmer, R. (2018). *Sprachentwicklungsstörungen bei frühen Zweitsprachlernern - der Erwerb von Kasus, Finitheit und Verbstellung [developmental language disorders in early second language learners - the acquisition of case, finiteness, and position of the verb]* [doctoral dissertation]. Goethe University Frankfurt, Frankfurt on the Main, Germany.

Lemmerth, N., and Hopp, H. (2018). Gender processing in simultaneous and successive bilingual children: cross-linguistic lexical and syntactic influences. *Lang. Acquis.* 26, 21–45. doi: 10.1080/10489223.2017.1391815

Lenneberg, E. H. (1967). *Biological Foundations of Language*. NJ: Wiley.

Lidzba, K., Konietzko, A., Schwilling, E., Krägeloh-Mann, I., and Winkler, S. (2013). Processing of non-canonical word-order: a case-series on lesion-induced reorganized language and age-effects in typical development. *Brain Lang.* 127, 377–387. doi: 10.1016/j.bandl.2013.10.010

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., and Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *J. Mem. Lang.* 69, 574–588. doi: 10.1016/j.jml.2013.08.001

Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-Native Speakers. Cambridge Applied Linguistics Series*. Cambridge: Cambridge University Press.

Meisel, J. M. (2018). Early child second language acquisition: French gender in German children. *Biling. Lang. Cogn.* 21, 656–673. doi: 10.1017/S1366728916000237

Mohamed, M. T. (2013). Accusative case-assignment in double object constructions in standard Arabic: a minimalist approach. *Asian Soc. Sci.* 10, 95–105. doi: 10.5539/ass.v10n1p95

Müller, G. (1999). Optimality, markedness, and word order in german. *Linguistics* 37, 777–818. doi: 10.1515/ling.37.5.777

Müller, N., and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Biling. Lang. Cogn.* 4, 1–21. doi: 10.1017/S1366728901000116

Naber, M., Frässle, S., Rutishauser, U., and Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *J. Vis.* 13:11. doi: 10.1167/13.2.11

Ortega, L. (2019). Sla and the study of equitable multilingualism. *Mod. Lang. J.* 103, 23–38. doi: 10.1111/modl.12525

Papesh, M. H., Goldinger, S. D., and Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *Int. J. Psychophysiol.* 83, 56–64. doi: 10.1016/j.ijpsycho.2011.10.002

Paradis, J. (2016). The development of english as a second language with and without specific language impairment: clinical implications. *J. Speech Lang. Hear. Res.* 59, 171–182. doi: 10.1044/2015_JSLHR-L-15-0008

Paradis, J. (2019). "English second language acquisition from early childhood to adulthood: the role of age, first language, cognitive, and input factors," in *Proceedings of the 43rd Annual Boston University Conference on Language Development*, eds M. M. Brown and B. Dailey (Bosto, MA: Cascadilla Press), 11–26.

Partala, T., and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *Int. J. Hum. Comput. Stud.* 59, 185–198. doi: 10.1016/S1071-5819(03)00017-X

Piquado, T., Isaacowitz, D., and Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* 47, 560–569. doi: 10.1111/j.1469-8986.2009.00947.x

Porter, G., Troscianko, T., and Gilchrist, I. D. (2007). Effort during visual search and counting: insights from pupillometry. *Q. J. Exp. Psychol.* 60, 211–229. doi: 10.1080/17470210600673818

Preuschoff, K., 't Hart, B. M., and Einhäuser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115

Primus, B. (2012). *Semantische Rollen [Semantic roles]*. Heidelberg: Universitätsverlag Winter.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed August 8, 2021).

Renner, L. F., and Włodarczak, M. (2017). "When a dog is a cat and how it changes your pupil size: pupil dilation in response to information mismatch," in *Interspeech 2017* (Stockholm).

Røreng, A. (2011). *Die deutsche Doppelobjektkonstruktion: Eine korpusbasierte Untersuchung zur relativen Abfolge nominaler Akkusativ- und Dativobjekte im geschriebenen Deutsch [the german double object construction: a corpus-based study of the relative sequence of nominal accusative and dative objects in written german]* [doctoral dissertation]. University of Tromsø, Tromsø, Norway.

RStudio Team (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, PBC. Available online at: http://www.rstudio.com/ (accessed August 8, 2021).

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., and Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: evidence from pupillometry and event-related fMRI. *NeuroImage* 37, 1017–1031. doi: 10.1016/j.neuroimage.2007.04.066

Sauerman, A., and Höhle, B. (2018). Word order in German child language and child-directed speech: a corpus analysis on the ordering of double objects in the german middlefield. *Glossa J. Gen. Linguist.* 3:57. doi: 10.5334/gjgl.281

Scheepers, C., and Crocker, M. W. (2004). "Constituent order priming from reading to listening: a visual-world study," in *The On-Line Study of Sentence Comprehension: Eyetracking, Erps, and Beyond*, eds M. Carreiras and C. Clifton (Hove: Psychology Press), 167–185.

Scheepers, C., Mohr, S., Fischer, M. H., and Roberts, A. M. (2013). Listening to limericks: a pupillometry investigation of perceivers' expectancy. *PLoS ONE* 8:e74986. doi: 10.1371/journal.pone.0074986

Scherger, A.-L. (2015). *Schnittstelle zwischen Mehrsprachigkeit und Sprachentwicklungsstörung: Kasuserwerb deutsch-italienischer Kinder mit spezifischer Sprachentwicklungsstörung [Interface between multilingualism and developmental language disorder: case acquisition of German-Italian children with specific language impairment]. Schriftenreihe Philologia: Vol. 200*. Hamburg: Verlag Dr. Kovač.

Scherger, A.-L. (2016). Kasuserwerb bilingual deutsch-italienischer Kinder: vorübergehend verzögernder Spracheneinfluss [Case acquisition of bilingual German-Italian children: Temporarily delaying cross-linguistic influence]. *Linguistische Berichte* 246, 195–239.

Scherger, A.-L. (2018). German dative case marking in monolingual and simultaneous bilingual children with and without sli. *J. Commun. Disord.* 75, 87–101. doi: 10.1016/j.jcomdis.2018.06.004

Scherger, A.-L. (2019). "Dative case marking in 2L1 and L2 bilingual SLI," in *Proceedings of Gala 2017: Language Acquisition and Development*, eds P. Guijarro Fuentes and C. Suaréz-Goméz (Cambridge: Cambridge Scholars Publishing), 95–115.

Scherger, A.-L. (2021). The role of age and timing in bilingual assessment: non-word repetition, subject-verb agreement and case marking in l1 and el2 children with and without sli. *Clin. Linguist. Phon.* doi: 10.1080/02699206.2021.1885497. [Epub ahead of print].

Schlenter, J. (2019). *Predictive language processing in late bilinguals: evidence from visual-world eye-tracking* [doctoral dissertation]. University of Potsdam, Potsdam, Germany.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain Lang.* 17, 133–145. doi: 10.1016/0093-934X(82)90010-4

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Front. Psychol.* 5:137. doi: 10.3389/fpsyg.2014.00137

Schmidtke, J. (2018). Pupillometry in linguistic research: an introduction and review for second language researchers. *Stud. Second Lang. Acquisit.* 40, 529–549. doi: 10.1017/S0272263117000195

Schmitz, K. (2006). "Indirect objects and dative case in monolingual German and bilingual german/romance language acquisition," in *Studies in Language Companion Series: v. 75. Datives and Other Cases: Between Argument Structure and Event Structure*, eds D. P. Hole, W. Abraham, and A. Meinunger (Amsterdam: J. Benjamins), 239–268.

Schmitz, K., Patuto, M., and Müller, N. (2012). The null-subject parameter at the interface between syntax and pragmatics: evidence from bilingual German–Italian, German–French, and Italian–French children. *First Lang.* 32, 205–238. doi: 10.1177/0142723711403880

Schönenberger, M., Sterner, F., and Ruberg, T. (2011). "The realization of indirect objects and dative case in german," in *11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, eds J. Herschensohn and D. Tanner (Somerville, MA).

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

Sirois, S., and Brisson, J. (2014). Pupillometry. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 679–692. doi: 10.1002/wcs.1323

Stölten, K., Abrahamsson, N., and Hyltenstam, K. (2014). Effects of age of learning on voice onset time: categorical perception of swedish stops by near-native L2 speakers. *Lang. Speech* 57, 425–450. doi: 10.1177/0023830913508760

Süss, A., Hendricks, P., Fritzsche, T., and Höhle, B. (2018). "Acquisition of adjectival agreement in german: sensitivity to grammar is reflected in 3-year-olds' pupil dilation," in *Bucld 42: Proceedings of the 42nd Annual Boston University Conference on Language Development*, eds A. Bertolini and M. Kaplan (Sommerville, MA), 722–735.

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., and Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *J. Exp. Child Psychol.* 153, 140–148. doi: 10.1016/j.jecp.2016.07.014

Ulrich, T., Penke, M., Berg, M., Lüdtke, U., and Motsch, H. J. (2016). Der Dativerwerb - Forschungsergebnisse und ihre therapeutischen Konsequenzen [Dative acquisition–research findings and their therapeutic consequences]. *Logos* 24, 176–190. doi: 10.7345/prolog-1603176

van Dijk, C., van Wonderen, E., Koutamanis, E., KOOTSTRA, G. J., Dijkstra, T., and Unsworth, S. (2021). Cross-linguistic influence in simultaneous and early sequential bilingual children: a meta-analysis. *J. Child Lang.* doi: 10.1017/S0305000921000337. [Epub ahead of print].

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., and Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends Hear.* 23:2331216519832483. doi: 10.1177/2331216519832483

Wechsler, D. (2017). *WISC-V: Wechsler Intelligence Scale for Children – Fifth Edition (F. Petermann, Trans.)*. London: Pearson.

Wilkinson, J. (2020). "From native speaker to intercultural speaker and beyond: intercultural (communicative) competence in foreign language education," in *Routledge Handbooks in Applied Linguistics. The Routledge Handbook of Language and Intercultural Communication*, ed J. Jackson (London: Routledge), 283–298.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R (2nd ed.). Texts in Statistical Science*. Boca Raton, FL: CRC Press.

Yu, A. J. (2012). Change is in the eye of the beholder. *Nat. Neurosci.* 15, 933–935. doi: 10.1038/nn.3150

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: current state of knowledge. *Trends Hear.* 22, 1–25. doi: 10.1177/2331216518777174

Zheng, W.-L., Dong, B.-N., and Lu, B.-L. (2014). "Multimodal emotion recognition using eeg and eye tracking data," in *36th Annual International Conference of the IEEE, Institute of Electrical and Electronics Engineers (Chair)* (Chicago, IL).

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

frontiers in Psychology

CORRECTION
published: 10 November 2021
doi: 10.3389/fpsyg.2021.790449

Corrigendum: The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry

Anna-Lena Scherger[1]*, *Gianna Urbanczik*[1], *Timon Ludwigs*[1] and *Jasmin M. Kizilirmak*[2,3]

[1] Department of Rehabilitation Sciences, Language & Communication, TU Dortmund University, Dortmund, Germany, [2] University of Hildesheim, Institute for Psychology, Neurodidactics & NeuroLab, Hildesheim, Germany, [3] German Center for Neurodegenerative Diseases, Cognitive Geriatric Psychiatry, Göttingen, Germany

Keywords: language production, language comprehension, ditransitives, implicit and explicit knowledge, article omission, pupillometry

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Anna-Lena Scherger
anna-lena.scherger@tu-dortmund.de

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

Received: *06 October 2021*
Accepted: *13 October 2021*
Published: *10 November 2021*

Citation:
Scherger A-L, Urbanczik G, Ludwigs T and Kizilirmak JM (2021) Corrigendum: The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry. Front. Psychol. 12:790449. doi: 10.3389/fpsyg.2021.790449

A Corrigendum on

The Bilingual Native Speaker Competence: Evidence From Explicit and Implicit Language Knowledge Using Elicited Production, Sentence-Picture Matching, and Pupillometry
by Scherger, A-L., Urbanczik, G., Ludwigs, T., and Kizilirmak, J. M. (2021). Front. Psychol. 12:717379. doi: 10.3389/fpsyg.2021.717379

In the published article, there was an error in affiliation 2. Instead of "University of Hildesheim, Institute for Psychology, Hildesheim, Germany", it should be "University of Hildesheim, Institute for Psychology, Neurodidactics & NeuroLab, Hildesheim, Germany."

Additionally, in the original article, there was an error in the Acknowledgments section. In the original version, it read "We thank Tom Fritzsche for helpful comments on the pupillometry experiment setup and Prof. Kristian Folta-Schoofs for allowing the usage of his lab."

A correction has been made to *Acknowledgments*. The corrected paragraph is shown below.

ACKNOWLEDGMENTS

We acknowledge financial support by Deutsche Forschungsgemeinschaft and TU Dortmund University within the funding programme Open Access Publishing. We acknowledge financial support by the University of Hildesheim for a student assistant. We thank Tom Fritzsche for helpful comments on the pupillometry experiment setup and Prof. Dr. Kristian Folta-Schoofs for allowing the usage of his lab and equipment. For their help with data collection, we thank Klara Tittelbach-Helmrich and David Mietzner. We further thank Hannes Elfers and the Elisabethschule in Hildesheim for their invaluable help with data collection and Julia Diedrich for illustrating the visual stimuli. Above all, we thank our participants.

Unfortunately, in the original article, a number of references also contained errors as listed below.

For Kauschke and Siegmüller (2010), the city of the publisher was incorrect. It was incorrectly written as Kauschke, C., and Siegmüller, J. (2010). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS) [Patholinguistic Assessment of Developmental Language Disorders in German], 2nd Edn*. Amsterdam: Elsevier. The corrected reference is shown below.

The reference for Preuschoff et al. (2011) contained an incorrect name for one author. It was incorrectly written as Preuschoff, K., Hart, B. M., and 't Einhäuser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115. The corrected reference is shown below.

The reference for Schmitz (2006) incorrectly contained a URL. It was incorrectly written as: Schmitz, K. (2006). "Indirect objects and dative case in monolingual German and bilingual german/romance language acquisition," *in Studies in Language Companion Series: v. 75. Datives and Other Cases: Between Argument Structure and Event Structure*, eds D. P. Hole, W. Abraham, and A. Meinunger (Ahttp://en.wikipedia.org/wiki/Amsterdam msterdam: J. Benjamins), 239–268. The corrected reference is shown below.

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Kauschke, C., and Siegmüller, J. (2010). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS) [Patholinguistic Assessment of Developmental Language Disorders in German], 2nd Edn*. Munich: Elsevier.

Preuschoff, K., 't Hart, B. M., and Einhäuser, W. (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front. Neurosci.* 5:115. doi: 10.3389/fnins.2011.00115

Schmitz, K. (2006). "Indirect objects and dative case in monolingual German and bilingual German/Romance language acquisition," *in Studies in Language Companion Series: v. 75. Datives and Other Cases: Between Argument Structure and Event Structure*, eds D. P. Hole, W. Abraham, and A. Meinunger (Amsterdam: J. Benjamins), 239–268.

frontiers in Psychology

POLICY AND PRACTICE REVIEWS
published: 30 September 2021
doi: 10.3389/fpsyg.2021.715843

The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures

*Lauretta S. P. Cheng[1], Danielle Burgess[1], Natasha Vernooij[2], Cecilia Solís-Barroso[1], Ashley McDermott[3] and Savithry Namboodiripad[1]**

[1]Department of Linguistics, University of Michigan, Ann Arbor, MI, United States, [2]Department of Psychology, University of Michigan, Ann Arbor, MI, United States, [3]Department of Anthropology, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez, University of the Balearic Islands, Spain

Reviewed by:
Helene Kreysa, Friedrich Schiller University Jena, Germany
Alba Tuninetti, Bilkent University, Turkey

***Correspondence:**
Savithry Namboodiripad
savithry@umich.edu

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

Received: *27 May 2021*
Accepted: *18 August 2021*
Published: *30 September 2021*

Citation:
Cheng LSP, Burgess D, Vernooij N, Solís-Barroso C, McDermott A and Namboodiripad S (2021) The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures. Front. Psychol. 12:715843. doi: 10.3389/fpsyg.2021.715843

Though the term NATIVE SPEAKER/SIGNER is frequently used in language research, it is inconsistently conceptualized. Factors, such as age, order, and context of acquisition, in addition to social/cultural identity, are often differentially conflated. While the ambiguity and harmful consequences of the term NATIVE SPEAKER have been problematized across disciplines, much of this literature attempts to repurpose the term in order to include and/or exclude certain populations. This paper problematizes NATIVE SPEAKER within psycholinguistics, arguing that the term is both unhelpful to rigorous theory construction and harmful to marginalized populations by reproducing normative assumptions about behavior, experience, and identity. We propose that language researchers avoid NATIVE SPEAKER altogether, and we suggest alternate ways of characterizing language experience/use. The vagueness of NATIVE SPEAKER can create problems in research design (e.g., through systematically excluding certain populations), recruitment (as participants' definitions might diverge from researchers'), and analysis (by distilling continuous factors into under-specified binary categories). This can result in barriers to cross-study comparison, which is particularly concerning for theory construction and replicability. From a research ethics perspective, it matters how participants are characterized and included: Excluding participants based on binary/essentialist conceptualizations of nativeness upholds deficit perspectives toward multilingualism and non-hegemonic modes of language acquisition. Finally, by implicitly assuming the existence of a critical period, NATIVE SPEAKER brings with it theoretical baggage which not all researchers may want to carry. Given the issues above and how 'nativeness' is racialized (particularly in European and North American contexts), we ask that researchers consider carefully whether exclusion of marginalized/minoritized populations is necessary or justified—particularly when NATIVE SPEAKER is used only as a way to achieve linguistic homogeneity. Instead, we urge psycholinguists to explicitly state the specific axes traditionally implied by NATIVENESS that they wish to target. We outline several of these (e.g., order of acquisition, allegiance, and comfort with providing intuitions) and give

examples of how to recruit and describe participants while eschewing NATIVE SPEAKER. Shifting away from harmful conventions, such as NATIVE SPEAKER, will not only improve research design and analysis, but also is one way we can co-create a more just and inclusive field.

Keywords: research methods, native speaker, psycholinguistics, language experience, multilingualism

INTRODUCTION

This article problematizes the use of NATIVE SPEAKER[1] as a construct in language research. We argue that the concept is both vague and harmful, and advocate for the field of psycholinguistics to move forward with a more careful, considered, and nuanced view of language experience.[2] We suggest that NATIVE SPEAKER is more accurately thought of as an ideology rather than an idealization, and give recommendations for how to shift our research practice accordingly.

The structure of the article is as follows. In Section "Introduction", we present background on what the problems are, with regard to both vagueness and harm. Next, we connect these broader concerns with why it is an issue within psycholinguistics research, specifically relating to methodology and theory. In Section "Assessment and Implications of Current Practices", we discuss implications of alternative approaches that could be taken in response to the big-picture issues detailed in Section "Introduction", drawing from the literature in other disciplines that already problematize the concept. This section is divided into three stages of the research process: conceptualization, recruitment, task, and survey design, and data analysis. In Section "Actionable Recommendations", we provide actionable recommendations for how psycholinguistic researchers can move away from NATIVE SPEAKER in their own work at each stage of the research process.

To represent a snapshot of the diversity of experiences that cannot be captured by NATIVE SPEAKER, we also provide four example profiles of language users that researchers may encounter (**Boxes 1–4**). We describe their language profiles (see the aspects of language experience laid out in "Complicating NATIVENESS in Recruitment, Tasks, and Surveys") and return to them as examples throughout the paper.

Vagueness

The term NATIVE SPEAKER is frequently used in language research. Perhaps precisely because of its frequent use and the assumption that it carries an intuitive meaning, the term is often not explicitly defined and operationalized. In some cases, where NATIVE SPEAKER is operationalized, the definition may be circular in nature and involve unstated, implicit assumptions. For example, according to Benmamoun et al. (2013), a NATIVE SPEAKER is someone that has "normal first language acquisition" (130) and that has "native" pronunciation. This description not only assumes that there are normal and abnormal acquisition processes (without detailing what those involve) but also refers to 'nativeness' in the definition to describe a criterion, ultimately failing to define the term NATIVE SPEAKER. Similarly, the antonym NON-NATIVE SPEAKER groups together an extremely heterogeneous set of individuals while strongly connoting a normative and monolingual experience. As Dewaele (2018b) puts it, NON-NATIVE SPEAKER is "inherently strange" as we are "defin[ing] somebody by what she or he is not" (236). If we take a closer look at the literature and compare definitions from various works, it is apparent that NATIVE SPEAKER is used more vaguely than one would imagine.

To exemplify the broad range of definitions for NATIVE SPEAKER, **Table 1** lists some definitions of this term as used across linguistics and adjacent fields. Certain concepts come up frequently, but while some definitions associate NATIVE SPEAKER with multiple factors (e.g., Benmamoun et al., 2013), others associate NATIVENESS with just one factor (e.g., Boltokova, 2017). Importantly, while there is an array of definitions for NATIVE SPEAKER, there are many more papers that use the term without defining it. Drawing on these and other examples, several common trends in usage can be identified across the literature: *nativeness-as-history* (including age, order, and context of acquisition), *nativeness-as-proficiency* (including continued usage), and *nativeness-as-identity*. We briefly demonstrate how each of these themes—alone or in combination—are realized in the context of various sub-fields of language research.

Linguistics as a field has historically conceptualized the NATIVE SPEAKER to be based on *proficiency* gained from a very specific "ideal" upbringing: They are the "ideal speaker-listener" with full mastery of a particular language and therefore able to provide authoritative judgments about grammaticality for any aspect of grammar (Chomsky, 1957, 1965). Chomsky's (1965) description of NATIVE SPEAKER also implies that a speaker's acquisition *history* must be monolingual[3] in

[1]Throughout this article, we use NATIVE SPEAKER in small caps to denote the construct. We use 'native speaker' in single quotes to refer to hypothetical uses (as in, the so-called native speaker). We reserve "native speaker" in double quotes for direct quotations. In all cases, we set the term NATIVE SPEAKER apart in order to signal that it is not used in a manner that assumes it as the default, unmarked state.

[2]Language *experience* is conceptualized broadly, encompassing all aspects of linguistic identity, proficiency, usage, input, output, language contact, etc. In comparison, we use language *history* to refer specifically to aspects of language development which are included in more "traditional" accounts of language acquisition, such as age, order, or context of acquisition. Crucially, we consider these factors to be subsumed under language experience.

[3]As the distinction between language and language variety is socially constructed, here and elsewhere, what we say about monolingualism and multilingualism could also apply to monolectalism (using one variety of a named language) and multilectalism (using multiple varieties of the same named language).

TABLE 1 | Some definitions of NATIVE SPEAKER.

Source	Definition of NATIVE SPEAKER	Facet(s) of language experience represented
Stern, 1983	"Native speakers have (a) a subconscious knowledge of rules, (b) an intuitive grasp of meanings, (c) the ability to communicate within social settings, (d) a range of language skills, and (e) creativity of language use." (154)	Proficiency
Abrahamsson and Hyltenstam, 2009	A native speaker "(a) has spoken only Swedish at home during childhood; (b) has had Swedish as the only language of instruction at school; and (c) has lived his or her whole life in a context in which Swedish has been the majority language" (264)	History
Debenport, 2011	Speakerhood-as-identity: "Tribal members who play significant religious or political roles are more likely to be counted by San Antonians as 'speakers'" (90)	Identity
Benmamoun et al., 2013	"A prototypical (educated) native speaker lives in a monolingual environment, or in a bilingual environment in which his/her original native language has not undergone attrition. Such a prototypical speaker is expected to have "native" pronunciation and a sizable, comprehensive vocabulary (about 20,000 words)" (130)	History and Proficiency
Rothman and Treffers-Daller, 2014	"A native language is one that is acquired from naturalistic exposure, in early childhood and in an authentic social context/speech community" (95)	History
	"Native speaker (i) bi-/multilinguals have multiple native languages; and (ii) nativeness can be applicable to a state of linguistic knowledge that is characterized by significant differences to the monolingual baseline."	

nature (such as the speaker profiled in **Box 1**), and this implication has been used as an underlying assumption in subsequent research (e.g., Sorace, 2004; Thráinsson, 2012; Benmamoun et al., 2013).

In line with this, researchers studying second language acquisition or bilingualism have been known to assume and employ NATIVE SPEAKER or LANGUAGE to mean a high (or the highest possible) degree of *proficiency*. In this way, it is commonly used as a benchmark or comparison group for language learners or those who are considered otherwise "non-native" (e.g., Au et al., 2002; Abrahamsson and Hyltenstam, 2009). However, for the purposes of operationalization, many studies rely on acquisition *history* to identify native speakers. For example, Abrahamsson and Hyltenstam (2009) recruited 'native speakers of Swedish' based on language history (see **Table 1**). While comparing "native" Swedish speakers to reportedly "native-like" but "non-native" Swedish speakers, they assert that nativeness is a binary phenomenon like "'marriedness' and 'deadness'" (267). The authors imply that there exists some threshold of language ability available to 'native speakers' that 'non-native speakers' cannot reach, as a result of age of language onset.

In other contexts, though 'nativeness' may still be taken for granted, the traditional definition of NATIVE SPEAKER breaks down. In sign language research, for example, the concept of a NATIVE SIGNER has proven to be rather elusive. This stems from the fact that signing individuals' experiences are highly heterogeneous and idiosyncratic, differing in a number of ways from normative spoken language experiences (Quer and Steinbach, 2019). The strict definition of NATIVE SIGNER, based on language *history*, would include only second-generation deaf signers—that is, deaf individuals growing up with (deaf) signing parents.[4] Since this is extremely uncommon, the idea of nativeness has been operationalized *via* a variety of criteria in sign language linguistics (e.g., Neidle et al., 2000; Costello et al., 2008; Mathur and Rathmann, 2009), involving aspects of *history*, *proficiency*, and *identity*. These have included some of the following: family environment (e.g., having deaf signing parents), early experience (e.g., prior to age 3), continued exposure and usage (e.g., daily contact), some indication of grammatical competence (e.g., ease of making judgments), and identification with the Deaf community.

In the context of immigration and language contact situations, continued usage (with links to *proficiency*) has also been implicated as a part of the definition of a NATIVE SPEAKER. Benmamoun et al. (2013) define "language attrition" as "the [gradual] loss of aspects of a native language by a healthy native speaker," going on to say that "a native speaker will become, in the judgment of his or her peers, a non-native speaker of his/her own language" (132). This suggests that the status of NATIVE SPEAKER in this conception relies on degree of language use and maintained language ability.

Finally, though much of the previously outlined usages of NATIVE SPEAKER/SIGNER appeals to, or at least brings with

[4]See discussion in Fisher et al. (2018) of how researchers' priorities in selecting Deaf participants who fit a particular profile may have influenced ideologies of belonging in certain (American) Deaf communities.

BOX 1 | An example profile of a primarily monolingual speaker.

Ingrid, 65, grew up in The Netherlands and attended Dutch-medium schools. She learned Dutch from Dutch-speaking parents and spoke only Dutch at home, with friends, and in society at large. She learned written English and German in school, from ages 8–15, but does not use those languages in her daily life. While she can read in both languages, she isn't comfortable speaking in either of them.

The written Dutch that she learned in school varies slightly from the spoken variety that she uses in her daily life at work. She is an avid reader of novels in Dutch and reads some news articles in English and German. She watches television and movies in Dutch and English, and uses Dutch subtitles for English media.

She considers herself to be a 'native' or 'mother tongue' speaker of Dutch.

BOX 2 | An example profile of a "heritage"/immigrant bilingual.

Amy, 23, was born in Hong Kong and lived there until age 2 when her family emigrated to Toronto, Canada. She was first exposed to and began speaking only Cantonese. After moving to Canada, she began to hear English *via* immersion in a preschool setting starting at age 3, but still spoke only Cantonese otherwise.

After starting elementary school around age 5, Amy began to spend increasingly more time exposed to English. She continued to hear and speak Cantonese at home with family (including some media like TV and songs), with a couple of family friends, at certain extracurriculars (e.g., Saturday school for Cantonese) and in some places in the community that she went to with family (e.g., church, restaurants, grocery stores). Otherwise, English was heard and spoken at school, with peers and friends outside of school, at most extracurricular activities (e.g., sports teams, volunteering/work) and most places in the community. During this time, she gradually became less comfortable using Cantonese to communicate.

Currently, Amy uses English for almost everything other than speaking to her parents. She considers herself to not be fluent in speaking or listening to Cantonese. Due to some years of Saturday school as a child, she can read and write a small amount. Based on this, she considers English as her dominant, strongest and effectively only language. She doesn't fully identify as a native speaker of either language, but she would say Cantonse is her mother tongue, while she speaks more like a (near-)native speaker of English.

it an assumption of, competence (cf., Dewaele, 2018b), not all definitions do. Rothman and Treffers-Daller (2014), for example, argue for the inclusion of heritage bilinguals (see example profile in **Box 2**) under the umbrella of NATIVE SPEAKERS (see **Table 1**). This definition is similar to the sign language research context in prioritizing early naturalistic exposure (*history*), but explicitly excludes the expectation of *proficiency* at any level.

Likewise, in contexts of language reclamation, being a (NATIVE) SPEAKER of a language does not come with connotations of proficiency at all, but instead is used in the sense of *identity* and membership. In the case of the Dene Tha community in Chateh, located in Alberta, Canada, among young people, "there is a strong self-identification with one's heritage language and culture and a deeply rooted personal belief in belonging, as full and rightful members, to this language community" (Boltokova, 2017, p. 22). For communities like this one (cf., Debenport, 2011 who discusses a Pueblo community in San Antonio), 'native speaker' only refers to identity. In this way, the Dene Tha youth consider themselves as 'native speakers' without necessarily speaking the language with high fluency—or as some would say, "native" proficiency.[5]

As we can see, there is no clear, consensus definition of NATIVE SPEAKER/SIGNER or NATIVE LANGUAGE. This multifaceted concept can, but does not always, involve a constellation of factors relating to age, order, and context of acquisition (e.g., Costello et al., 2008; Abrahamsson and Hyltenstam, 2009; Rothman and Treffers-Daller, 2014; Quer and Steinbach, 2019), continued usage and/or exposure (e.g., Costello et al., 2008; Benmamoun et al., 2013), proficiency or competence (e.g., Chomsky, 1965; Abrahamsson and Hyltenstam, 2009), and sociocultural identification or membership (e.g., Debenport, 2011; Benmamoun et al., 2013; Boltokova, 2017).

Not only do researchers include different combinations of the above factors in their definition, the specifications of each criterion can also vary. Proficiency, for instance, is sometimes assumed, sometimes measured in a certain domain *via* assessments or tasks, and sometimes related to dominance, where the "strongest" language is considered the native language, rather than some measure of "absolute" level of proficiency. Age of acquisition, while often invoked, varies as to the exact ages that matter, for example, birth (Johnson and Newport, 1989) or age 3 (e.g., Costello et al., 2008; Dewaele, 2018b). Throughout many of these uses, the term is associated with assumptions of monolingualism and acquisition as a first language in contexts where there are clear temporal orders to learning different languages (Cook, 1999; Benmamoun et al., 2013; Dewaele, 2018b). All this together suggests that NATIVE SPEAKER, when used, can and does refer to disparate aspects of language experience across fields, studies, and contexts. We argue that the vagueness of this term is one reason to reconsider the extent to which NATIVE SPEAKER/SIGNER is a relevant and useful concept in our research.

Harm

All conceptual categories used for research are inherently simplifications and cannot capture the complexity of social life, but are necessary because there is no way to conduct meaningful research without them. However, a sensible scientific aim can be to ensure that the terms we use both describe the phenomenon of interest as accurately as possible and do not harm the communities we study. The term NATIVE SPEAKER meets neither requirement: (i) As argued in the previous section, it is ambiguous and thus a hindrance to data analysis and rigorous theory construction, and (ii) as we argue in this section, it can be harmful, particularly to minoritized individuals and groups, in that use of NATIVE SPEAKER in academic research reproduces normative assumptions about linguistic behavior, experience, and identity.

As illustrated in Section "Vagueness", the range of its use in research implies that the NATIVE SPEAKER is an "ideal speaker listener" (Chomsky, 1965) who has had a particular acquisition experience (learning one named language in childhood in a linguistically homogeneous environment before

[5]Here, we find it relevant to highlight another way that the term NATIVE is commonly used, that is, to refer to the First Nations of North America. This adds another facet of ambiguity in how this term might be used and interpreted.

learning other languages); who is "highly proficient" in one named language; who has continued to use the same named language from childhood to adulthood; and for whom that language is part of their sociocultural identity. Crucially, this does not take into account the fact that a single named language cannot fulfill all of these roles for most individuals and communities around the world due to structural factors, such as globalization, colonialism, ableism, and linguistic discrimination of various types (cf. Ortega, 2020).

These normative assumptions of NATIVE SPEAKER reinforce hegemonic conceptions of language use, ability, acquisition, and linguistic identity. When researchers use NATIVE SPEAKER in their work, and when participants are excluded from research because they do not fit researcher expectations of a NATIVE SPEAKER, they perpetuate deficit perspectives toward multilingualism and non-hegemonic modes of language acquisition. This can (perhaps, inadvertently) frame these individuals and their practices as deviating from the norm, thus contributing to racialized conceptions of "nativeness" and feelings of LANGUAGELESSNESS among those whose speech is positioned as abnormal, in which individuals might be categorized (either by themselves or others) as not speaking any language at all (Rosa, 2016; Ramjattan, 2019). The linguistic experience of most of humanity does not conform to these assumptions, making the term both widely inapplicable and harmful, as it leads to the systematic exclusion of marginalized populations and perpetuation of deficit perspectives.

To return to an example from Section "Vagueness", most conceptions of NATIVE SPEAKER exclude the overwhelming majority of signers from being considered as "native" for research purposes. According to Quer and Steinbach (2019), most deaf children are not raised in environments "where there is *adequate sign language input* for the child to develop language competence *in a natural way*" and "do not fall under the strict definition of native speakers or signers" (2, emphasis our own). Limiting sign language research to only include deaf children born to deaf adults would not be representative of the use of signed languages in the world (see example in **Box 3**), and it is moreover harmful to position the acquisition contexts of the majority of signers as being inadequate, especially without attending to the structural reasons for this (i.e., Oralism and other forms of ableism).

Bucholtz (2003) discusses how the concern for "real language" and "authenticity" has made monolingualism appear unmarked in sociolinguistics. In the ideology of linguistic isolationism, research is based on the assumption that "the most authentic speaker belongs to a well-defined, static, and relatively homogenous social grouping that is closed to the outside" (404) and that "bilingualism and multilingualism are […] special rather than typical sociolinguistic situations" (405). Again, by positioning these linguistic experiences as abnormal and inauthentic, such frames position individuals themselves as abnormal and inauthentic.

In English Language Teaching (ELT), NATIVE SPEAKERISM and its associated harm have been deeply theorized, with Holliday (2006) arguing that this term represents an ideology that 'native speakers' are better equipped to teach English than 'non-native speakers.' Scholars who have expanded upon this work have shown that who is seen as a NATIVE SPEAKER of English is racialized, and prizing the speech and labor of perceived NATIVE SPEAKERS of English also ends up prizing whiteness (Gerald, 2020). In the context of Canada, Ramjattan (2019) shows how White 'native speakers' are perceived as being better teachers and more qualified, and, even beyond the context of North America, White speakers of English are more likely to be perceived as "native" (Sung, 2011; Lee and Jenks, 2019).

BOX 3 | An example profile of a mobile Deaf signer.

Angel, 38, is deaf and was born in Manila, Philippines. She moved to San Francisco, CA, in her 30s and now lives in Boston, MA, with her wife, who is also deaf. Growing up, she spoke/signed Tagalog, English, and Filipino Sign Language at home. In school, she used Tagalog and English, as she was integrated into a class of hearing students ("mainstreamed") on her own, with no interpreters or special support. Now, she primarily uses English and American Sign Language (ASL), and considers English to be her strongest language.

Angel attended school from the age of 5; in the 1980s and 1990s there was no organized educational interpreting system in the Philippines, so she was immersed in a spoken language environment. Some of her family is deaf, and they primarily communicate in Filipino Sign Language. At home, her family also used Tagalog, English, Bisaya, and Hokkien. Angel took courses in Nihongo (Japanese) at a language institute and took online courses in Japanese Sign Language, which she signs with a few friends. In college, she took a Castilian Spanish course. She learned ASL from her wife and YouTube videos, as well as from interactions in Deaf spaces in the United States.

In informal settings, Angel is most comfortable speaking Taglish (code-mixing of Tagalog and English), followed by ASL. ASL is the language that she uses the most with her wife, kids, friends, and coworkers at the university where she works. At home, she uses ASL, English, and Tagalog. At work/school, she uses ASL and English. She uses ASL with her friends. Angel uses English with strangers, but uses ASL if the stranger happens to know it. She considers herself a native speaker/signer of Tagalog, English, and Filipino Sign Language.

While many studies connecting race and nativeness are situated in the context of ELT, the ideologies that are described are certainly not limited to these contexts. Rosa and Flores (2017) discuss how "unaccented English" is conceptualized by English users as an English which conforms to White listeners' expectations. Rubin (1992) and following studies (Babel and Russell, 2015; Kutlu, 2020; et alia) using a matched guise paradigm have shown repeatedly that recordings played alongside White faces are rated as "more native," "more intelligible," or "less accented" than non-White faces. These ideologies are present in the world, and there is an opportunity for language research practitioners who want to create a more inclusive discipline to denaturalize the often implicitly made connections between 'nativeness' and race. This can be done by accounting for the possibility that our participants may hold these ideologies, as well as accounting for the possibility that we as researchers may also hold these ideologies, which can lead to systematic exclusion of racialized individuals from research.

Monolingualism as the norm is often implied in the term NATIVE SPEAKER (see "Vagueness"), and such assumptions

have harmful consequences for multilingual (or multilectal) individuals, especially those who are racialized. Rosa (2016) discusses how deficit perspectives are employed when talking about the natural multilingual practices of racialized individuals, such as language mixing. Rosa shows how languagelessness is assigned to racialized individuals who enact non-normative language practices; these individuals are labeled as not speaking any language at all. This feeling of languagelessness is also ascribed by multilingual individuals to themselves and is certainly not unique to the Global North; in the context of South India, Namboodiripad (2021) showed that Malayalam speakers for whom Malayalam was their first and most-used language felt that, because they mixed languages, they were not able to speak any language at all (see also **Box 4**). Across contexts, not conforming to a monoglot norm can make speakers themselves feel deficient, and institutional sites of language evaluation, whether in schools or in psycholinguistics experiments, can reinforce these negative and harmful ideologies.

The harm in creating a context in which languagelessness is imputed to (particularly racialized) multilinguals must be understood in a historical context in which ascribing languagelessness has been a tool for dehumanization. Degraff (2005) discusses how delegitimizing the languagehood of creoles was used as a tool of colonialism and chattel slavery, and how it led to dehumanizing the people who were being oppressed. DeGraff points out the continuity of this dehumanization in how creoles are described as exceptional languages that are birthed from "imperfect learning," and connects this to colonial narratives about how creoles were the result of "a race that is linguistically inferior" trying to learn colonial languages (Vinson, 1889, cited in Degraff, 2005). Looking at discourses on African American Language(s), we see similar deficit perspectives which are based on essentialist ideas about language attainment and learning (e.g., Green, 2004 on "dual components" approaches). Constructs, such as NATIVE SPEAKER, carry with them essentialist and harmful ideas about language and linguistic attainment, and psycholinguists who would like to push against such harm should reject and work against narratives which dehumanize our participants, our colleagues, and ourselves.

BOX 4 | An example profile of a multilingual individual experiencing globalization.

Leela, 27, was born in Kerala, India. The first language she was exposed to and spoke was a high contact variety of Malayalam, including elements from Tamil, Hindi, and English. She attended English immersion school, starting at age 4. While she learned to read and write Malayalam in school, Malayalam language classes ended at age 12. She also learned to read and write Hindi in school from ages 8–12.

Growing up, she heard and spoke Malayalam mostly at home with her family and in the community at large. She also heard and spoke English and Malayalam in school with peers and friends. She consumed mostly Malayalam and English media, but also sometimes watched Hindi and Tamil movies. Currently, she uses a high-contact variety of Malayalam at home, in social settings, and with greater society. She uses Malayalam and English at work (English with clients, superiors, and for all written communications; Malayalam with friends and in casual conversations); English when traveling outside of India; and English and Hindi when visiting family in North India. She also watches Hindi, Malayalam, English, and Tamil movies.

She doesn't consider herself fluent in Malayalam both because she prefers to read and write in English (though she can read and write in Malayalam), and because she doesn't feel like she can speak Malayalam without using English elements, especially depending on the semantic domain. She considers Malayalam her Mother Tongue, which is a locally relevant term, but states that she doesn't feel proficient in any language.

Connections to Research Methods and Theory

A well-designed experiment is detailed, makes clear predictions, targets a specific population, and involves a data analysis plan. Given this, the vague and harmful definitions of NATIVE SPEAKER pose significant methodological problems for the field of psycholinguistics.[6] As evidenced in the sections above, researchers make various implicit assumptions about the language experience of a NATIVE SPEAKER and these assumptions shape all aspects of research, ranging from question creation to data analysis. We make explicit these connections in this section, focusing on research conceptualization and design followed by comparison groups and analysis. As we will see, using NATIVE SPEAKER can lead to imprecise predictions, ill-selected samples, exclusionary and inconsistently defined participant pools, inappropriate materials, and misguided analyses.

Issues With NATIVE SPEAKER at the Stages of Conceptualization and Design

Some psycholinguistic research deals with specific predictions about how "native" and "non-native" speakers process or produce language (e.g., production of phonetic variability; Baese-Berk and Morrill, 2015; Vaughn et al., 2019; see also Bosker et al., 2014 on hesitation phenomena) or how different listeners process or perceive "native" or "non-native" (-sounding) language (e.g., comprehension, perceptual adaptation, or credibility of "foreign-accented" speech; Lev-Ari and Keysar, 2010; Hanulíková et al., 2012; Baese-Berk et al., 2013; Lev-Ari, 2015; Bent et al., 2016). However, even when NATIVENESS is not central to their research questions, researchers tend to recruit a sample of 'native speakers' for participation or for stimuli development. From a superficial search of research in top psycholinguistics journals, we found some examples which illustrate how this often looks in psycholinguistics research. Here is one pair of examples which covers participants: "Thirty native speakers of English from the University of York student community took part in this study" (Altmann, 2004), and "Twenty-four University of Rochester undergraduates who were native speakers of American English … were paid $10" (Kurumada et al., 2014). In the same papers, we found examples of how those selected to create stimuli are often described: "The sentences were recorded by a male native speaker of British English"

[6]For the purposes of this paper, we are construing psycholinguistics broadly, to include those who research sentence processing, speech perception, language development, and any other research area in which experimental methods might be used to investigate language perception, production, learning, and/or comprehension.

(Altmann, 2004), and "A native speaker of American English recorded two tokens of each item" (Kurumada et al., 2014).[7]

As illustrated in Section "Vagueness", NATIVE SPEAKER varies in definition across researchers and often goes completely undefined. One problem with this is that the researcher's idea of who a NATIVE SPEAKER is may not match the participant's idea of who a NATIVE SPEAKER is, if the concept is even clear to the participant. Faez (2011), for instance, presents several case studies that illustrate how (i) participants can have difficulty answering the question of whether they are "native" and (ii) an individual's self-ascribed "native" status may not be matched by the judgments of outside informants (nor do judges always agree with each other). This can lead to a discrepancy between the researcher's target sample and the actual individuals recruited. Another consequence is that different laboratories may be using the term inconsistently, which is problematic for replications, follow-up studies, and cross-study comparisons. For example, NATIVE SPEAKER study inclusion criteria commonly exclude the so-called "heritage speakers," such as Amy (**Box 2**), from participation—often based on harmful deficit perspectives—but some do not. Moreover, the extent of social and linguistic variation across those who are identified as 'native speakers' may not be fully considered by researchers. In this case, the inclusion of imprecisely defined 'native speakers' in stimuli norming or creation may result in biased stimuli, or stimuli that are inappropriate for the target participant demographic and research question (e.g., the regional variety spoken by a speaker recording audio stimuli or judging acceptability for stimuli may differ compared to the participants in the experiment).

Simply reporting that 'native speakers' participated or recorded stimuli clearly does not provide information adequate for replication. These issues are especially concerning given the replication crisis plaguing psychological research (Open Science Collaboration, 2015; Camerer et al., 2018). Henrich et al. (2010) noted that narrow samples from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations are frequently used to make broad claims about human psychology and behavior.[8] So, too, linguists have traditionally used monolingual speakers from relatively homogenous (and WEIRD) speech communities as a baseline to make broad claims about language organization and behavior (Evans and Levinson, 2009; Croft, 2013; Dahl, 2015; see also Sedarous and Namboodiripad, 2021 on the overrepresentation of Written, Institutionally supported, Standardized, and Prestigious (WISPy) languages in psycholinguistics). It is important to recognize that such speakers are exceptional, rather than the default case. The centering of binary, essentialist conceptualizations of native language competence leads to the exclusion of minoritized communities from research, as described in Section "Harm".

In addition, the implicit assumptions underlying NATIVE SPEAKER inherently espouse certain theoretical frameworks. For example, many researchers cite that learning a language at a young age is a key aspect of what defines a native speaker (e.g., Cook, 1999; Costello et al., 2008; Rothman and Treffers-Daller, 2014; Hall et al., 2017) and anyone who learns a language after this arbitrary cutoff (see discussion in "Vagueness") is no longer a NATIVE SPEAKER. While the existence of a biological critical (or sensitive) period is debated and not all language researchers agree with the idea (see, for example, Schouten, 2009; Balari and Lorenzo, 2015), this is precisely the theory that researchers implicitly adopt when they use NATIVE SPEAKER as a proxy for "someone who learned a language at a young age." For researchers who do not agree with the idea of a critical period, invoking this theoretical framework when they use NATIVE SPEAKER can be an issue. Researchers should be aware that the way that they define NATIVE SPEAKER may invoke theoretical frameworks that they do not necessarily agree with.

Inconsistent and vague definitions of NATIVE SPEAKER also pose issues for theory construction. By building solely upon conclusions from studies that inconsistently define their variables, use differing methods to categorize participants, and analyze participants based on these groupings (see "Issues with NATIVE SPEAKER When Constructing Comparison Groups and Conducting Analyses"), we will invariably create theories that fall prey to these same issues. In addition, the hegemony of research conducted in contexts where monolingualism/normative language use is seen as a control or neutral mode structures our fields of inquiry such that multilingualism/non-normative language use is peripheralized, requiring extra theoretical and methodological machinery. This privileges and incentivizes the study of certain, dominant groups over others, and puts research on socially less-powerful groups at an inherent disadvantage. Taking NATIVE SPEAKERS out of the center of our fields will not only sharpen research questions, but also expand the types of research that is done.

Issues With NATIVE SPEAKER When Constructing Comparison Groups and Conducting Analyses

When operationalizing concepts, psycholinguists often categorize participants into groups for the purposes of comparison, either *a priori* or in *post-hoc* examination of the collected data. Groups can include NATIVE SPEAKER, NON-NATIVE SPEAKER, L2 LEARNER, HERITAGE SPEAKER, and countless others. While many researchers use dimensions of language experience, such as age of acquisition, order of acquisition, or continued exposure, often collected *via* language experience questionnaires, others use proficiency tasks to group participants (e.g., picture naming, cloze tasks, and standardized language tests).

These different methods and criteria lead to huge variation in how participants are categorized and treated in analyses

[7]n.b., We are not intending to call out any particular authors with these examples—they are illustrative of the norms in psycholinguistics as we see them.
[8]See also Rochat (2010)who argues that psychologists need to go beyond simply diversifying participant populations and points out that WEIRD is itself a binary category which does not capture all the facets of overrepresentation. In addition, Clancy and Davis (2019) point out that WEIRD is racialized, in that it almost always means white, which adds another important dimension to which types of participants are systematically included and excluded.

across studies. While the use of standardized measures or tasks may appear more objective or consistent, different assessments can often categorize participants in significantly different ways (Solís-Barroso and Stefanich, 2019). This poses a crucial question for the field of psycholinguistics: Are the groups in language experiments comparable across studies? Further, researchers' constructed categories contain their assumptions about the term (NON-)NATIVE SPEAKER and ignore the ways in which these artificially "different" groups could be similar in their perception and production of language (for some work that problematizes this assumption, see Dąbrowska, 2013; Han et al., 2016; Johns et al., 2018). These categories are used in analyses to make broad generalizations about diverse and ill-defined groups of speakers which may not apply across different segments of the "same" population.

Of course, individual differences will always exist in any sample. However, when using underspecified categories, such as (NON-)NATIVE SPEAKER, there may also exist large systematic "within-group" differences that can affect the linguistic variables of interest, and therefore complicate theoretical interpretation and generalization. Because we often assume homogeneity in such groups, potentially relevant factors of language experience are inconsistently reported. For example, Surrain and Luk (2019) found in their review of research comparing bilinguals to monolinguals that there were systematic regional differences in what types of information was collected, with sociolinguistic context being about three times as likely to be reported when taking place outside of North America and Europe. If these types of linguistic experience are not collected or reported, we may not know the extent to which different samples are comparable.

To take one example, MONOLINGUAL NATIVE SPEAKERS in fact vary immensely in experience with other languages or varieties; though researchers do not always take this into account, it can lead to significant differences in linguistic behavior. For instance, lifetime experience with other speech varieties (e.g., living in an urban metropolis vs. small rural town) influences comprehension (Laturnus, 2018), while even short periods of exposure to another language can affect "native language" speech production (e.g., Chang, 2012). In addition, Dewaele (2018a) discusses how exposure to British English caused semantic restructuring in individuals who had grown up using American English. It behooves us to remember that the difference between a language and a variety is gradient and socially constructed, and, as such, even "Native Speakers of American English" represent a highly heterogeneous population whose particular language histories are likely relevant for various psycholinguistic processes.

These issues relate to a noted historical tendency of "categorical thinking" in psychology and psycholinguistics *via* an overreliance on factorial design and treating continuous predictors as categorical in analysis. The risks of discretizing continuous measures for analysis and the benefit—or indeed, necessity—of maintaining these continuous measures have also been argued for language research on conceptual and empirical grounds (see MacCallum et al., 2002; Baayen, 2004; Balota et al., 2004; Young, 2016). Categorizing continuous variables can not only lead to a decrease in statistical power, increased potential for spurious finding, and a reduced ability to detect complex and/or non-linear relationships (see Cohen, 1983; Young, 2016), but also have more general implications for interpretation and theory formation.

These problems are directly relevant to the case of experimental linguistic research where measurement of 'nativeness' and associated concepts (e.g., bilingualism and language dominance) has been historically inconsistent across the literature and often relies on binary categorization of continuous variables (e.g., Solís-Barroso and Stefanich, 2019; Ortega, 2020). Solís-Barroso and Stefanich (2019) show through the comparison of different measures of language dominance that treating language dominance as a categorical variable is problematic, given that an individual bilingual will not be consistently placed into the same dominance group depending on which assessment is given, contributing to potential heterogeneity within each group. By moving away from categorical thinking when it comes to participants, we allow for the discovery of more precise factors which influence language understanding and use.

ASSESSMENT AND IMPLICATIONS OF CURRENT PRACTICES

In this section, we discuss possible responses to the problems outlined in Section "Introduction", drawing from the approaches taken by researchers from various fields. These are organized into three (broadly defined) stages of research: conceptualization, recruitment, task, and survey design, and data analysis. We review the effectiveness of these solutions, along with representative examples, leading into our more specific recommendations in Section "Actionable Recommendations".

Complicating NATIVENESS in Conceptualization

Although scholars have suggested both narrowing (e.g., Cook, 1999) or broadening (e.g., Rothman and Treffers-Daller, 2014) the scope of the definition of NATIVE SPEAKER, others have argued to leave behind (NON-)NATIVE SPEAKER (and "mother tongue") in favor of more specified characterizations. For example, Rampton (1990) recommended decomposing NATIVE SPEAKER into (i) language *expertise* (linguistic knowledge and ability) and (ii) language *loyalty/allegiance* (social identification which can both be gained *via inheritance* or *affiliation*). These alternatives allow us to conceptualize two different facets (analogous to *proficiency* and *identity*) without appealing to NATIVENESS. For example, Amy (**Box 2**), a "heritage" bilingual, might consider herself an *expert* in English with less *expertise* in Cantonese. At the same time, she holds *allegiance* to both Cantonese (*via inheritance*) and English (*via affiliation*). This distinction on its own, however, does not specifically account for aspects of linguistic *history*.

Dewaele (2018b) proposed that we replace NATIVE and NON-NATIVE SPEAKER labels with L1 and LX (e.g., L2,

L3, and L4) in an effort to more specifically represent language history (i.e., early experience) separate from aspects of proficiency and identity. Thus, a researcher with specific hypotheses about individuals who learned a language earlier or later in life could compare them without drawing on NATIVENESS. For Amy (**Box 2**), Cantonese would be her L1, while English could be considered an LX—however, Dewaele's proposed cutoff for counting a language as LX is "after the age of 3 years" which may designate English as a second L1 for Amy. Nevertheless, while this terminology gets away from the hegemonic associations of NATIVE vs. NON-NATIVE, it still contains assumptions of normative ordered acquisition, which is not the case in many multilingual or globalized communities, and a critical period effect, which carries with it many other theory-specific assumptions. For Angel (**Box 3**), there are several possible L1s, but many of them are not very relevant for her language use across the bulk of her lifespan. These examples demonstrate how labels, such as L1 and LX, require researchers to rely on categories that may not be well-motivated, a practice that comes with many disadvantages (see "Complicating NATIVENESS in Data Processing and Analysis" and "Alternatives to NATIVE SPEAKER in Data Processing and Analysis" on Continuous Variables).

Regardless of the alternative terminology one chooses to use as labels or descriptors, we believe the best practice is to use specific characterizations of particular aspects of language experience (e.g., proficiency, history, and identity). This aligns with recommendations within bilingualism research to increase comparability across laboratories and studies by "provid[ing] detailed descriptions of the populations tested following a consistent approach" (Marian and Hayakawa, 2021, p. 7). By avoiding the conceptualization of (NON-)NATIVE SPEAKER at all levels of research, we are able to simultaneously (i) clarify our theoretical stance and interpretations, (ii) reject normative assumptions of the background of (NON-)NATIVE SPEAKERS, and (iii) acknowledge the heterogeneity of linguistic knowledge and behavior, even among those with purportedly similar backgrounds.

As discussed in Section "Issues With NATIVE SPEAKER When Constructing Comparison Groups and Conducting Analyses", even within supposedly MONOLINGUAL NATIVE SPEAKER populations from the same region, literacy and education can vary extensively, not to mention diversity in experience with other languages or language varieties. For example, Ingrid (**Box 1**) is relatively well-read and thus has consistent exposure to the types of syntactic structures disproportionately represented in written Dutch, such as sentences with multiple embeddings (see van der Wouden et al., 2002). However, another self-identified Dutch monolingual may not read much literature and therefore have a meaningfully different amount of exposure to multiple-embedded constructions. These factors, if not taken into consideration by the researcher, could lead to groups that are overly heterogeneous or otherwise not well-controlled. Clearly, delineating the relevant and irrelevant characteristics for the target research sample would allow for better control over homogeneity (if homogeneity is, indeed, the goal of the researcher—see "Harm" and "Issues With NATIVE SPEAKER at the Stages of Conceptualization and Design" for reasons why this might not be a desirable goal), as well as ensuring that groups used for purposes of comparison are indeed comparable or contrastive on the dimensions of interest to the researcher.

An additional benefit is that this practice, through increasing deliberate and thoughtful development of inclusion and exclusion criteria, can help to minimize the exclusion of underrepresented groups or individuals who do not fit into normative assumptions, but may in fact match the criteria for a particular research sample. This also potentially expands our participant pool, which may provide the practical advantage of facilitating research recruitment. In a similar vein, although a common practice in psycholinguistic research is the recruitment of 'native speakers' to norm or judge stimuli, or 'native, monolingual speakers' to record auditory stimuli, we cannot assume that these individuals have "neutral" identities or normative language histories. The same careful consideration of relevant and irrelevant characteristics of individuals under study should be applied to perceptual judges and speakers for stimuli.

Complicating NATIVENESS in Recruitment, Tasks, and Surveys

As researchers who have employed a critical approach to understanding research methods in linguistics and related fields have shown, it is imperative to understand how the questions we ask and the ways we ask them will be interpreted in local contexts, because these may differ significantly from researchers' expectations (Briggs, 1986; Hill, 2006). All research with human subjects, even research on language itself, constitutes a communicative event between research participants and researchers. Coming from a background in the academy, we have certain expectations about the communicative event that constitutes research, and the accepted norms of this event. However, our participants often do not come from a background of institutionalized research agendas and have differing familiarities with the communicative routines used (from survey, interview, and elicitation, to experimental task) and therefore do not share the same expectations for how the communicative event should unfold. Particularly, using the term "native" in recruitment may be understood differently by participants depending on their backgrounds, the context, and their understandings of research. In addition, translations of the term "native" could be interpreted in ways that researchers did not intend when designing the study, and unintentionally include or exclude participants.

As an example, a study of Hindi-Urdu sentence acceptability (Upreti and Namboodiripad, in prep.) asked a range of questions about exposure, comfort, and use of Hindi-Urdu. They also asked participants, at the end of the language experience survey, to indicate if they consider themselves "native speakers" of (a) Hindi-Urdu and (b) English. There were several cases in which participants' self-identification as "native speaker" did not align with many commonly understood correlates of nativeness: One participant had grown up in Pakistan (where

Urdu is an official language) and spent the first 35 years of his life in regions where Hindi-Urdu were the dominant languages. English and Hindi-Urdu were among the languages he heard growing up, and these were the two languages of instruction in his schooling. He rated himself as being maximally comfortable in reading, writing, listening to, and speaking Hindi-Urdu. And yet, he selected "no" when asked if he considered himself a native speaker of Hindi-Urdu (he also selected "no" for English, the language he considers his strongest). If that question had been the one used to recruit participants, this person would not have opted in, despite the fact that his language experience, as measured by the other questions, is highly relevant for this study.

Conceptions of the term NATIVE SPEAKER vary widely depending on how it is translated and the local context of its use. In many cases, some variant of "native language" is used on government censuses, and understandings of the term are influenced by historical uses and its connection to this survey. For example, in post-Soviet states, Soviet language planning tied languages to national territories, called Soviet Socialist Republics (SSRs), within the USSR. Language became a salient marker of national identity linked to the ability to find employment and access to resources in SSRs, and it was beneficial for people to identify their native language as that of the titular nationality (Slezkine, 1994; Martin, 2017). In the present, ties between native language and nationality still exist, and thus, asking for someone's "native language" will most likely elicit a response coinciding with their identity (e.g., Kyrgyz language, Kyrgyz, Kyrgyzstan) even if they speak primarily Russian, feel most comfortable speaking Russian, and learned Russian in early childhood. Thus, it is important during recruitment to ask tailored, specific questions about language use and experience to recruit appropriate participants, rather than asking whether they are a NATIVE SPEAKER of a language.

We also must consider that translations are not one-to-one equivalents of meaning, and the categories of NATIVE SPEAKER and its translations, though they will most likely overlap somewhat with understandings of the term in English, will not match completely. The common translation of "native language" in Russian, "rodnoĭ iazik [родной язык]," comes from the root "rod [род]" with associations of "birth" or "tribe" or "natural" and is found in words, like parents (roditeli [родители]) and homeland/motherland (rodina [родина]; Patrick, 1989). A rough translation is more like "mother tongue" or "birth tongue," but also carries connotations of national identity from its association with states (motherland) and historic uses in the census and language planning described above. As a consequence, a study asking for participants who are "native speakers of Russian" might recruit people who are from Russia, who identify as Russian, and whose parents or family speaks Russian but leave out those who believe Russian is the language they are most proficient in. Most problematically, asking this question could exclude the millions of non-Russians who use Russian, and misrepresent Russian as it is spoken by the majority of Russian-speakers in the world. Asking more targeted questions about the aspect of language use that is pertinent to the study, such as those outlined in **Table 2**, and understanding how these questions might be interpreted in local contexts will mitigate problems arising from unequivalent translations and conceptions of the term.

Given that people have different understandings of the term, particularly across sociolinguistic contexts, we note that merely asking participants to report whether they are 'native speakers'—without explicitly stating how the term is defined—is also not effective for assessment. Additionally, multilingual speakers often tend to under-rate or be unsure of their own linguistic abilities as a result of negative discourse surrounding their speech communities. This may influence identity as a 'native speaker' and is especially relevant for multilinguals who might be racialized (Tomoschuk et al., 2019; Ortega, 2020; see also Gullifer et al., 2021 on context-based mismatches between "native" language proficiency self-ratings and objective task measures). We see these issues reflected in the example profiles (**Boxes 1–4**) where individuals' self-characterization of their native language(s) varies and may not always align with researchers' goals (see also case studies in Faez, 2011).

In a study comparing four language dominance measures, some of which are also used to test 'nativeness,' Solís-Barroso and Stefanich (2019) found that out of 29 Spanish/English bilinguals tested, 20 were categorized differently depending on which measure was used. That is, even if multiple assessments claim to measure the same factor (e.g., 'nativeness,' proficiency, or dominance), they do not always yield the same categorization of a participant. To illustrate, imagine that Measure A defines 'nativeness' as solely (1) being born in a household where that language is spoken. Separately, Measure B operationalizes 'nativeness' as (2) being highly proficient in a language and (3) having no detectable 'accent.' Finally, Measure C requires all three criteria be met to be considered "native." In different studies, then, Amy (**Box 2**) could be considered "native" in Cantonese only (Measure A), "native" in English only (Measure B) or neither "native" in Cantonese nor English (Measure C). Note that these outcomes would not necessarily align with self-report either, which also may vary depending on whether she is asked about her first language, "mother tongue" or "native language."

Further, it is important to highlight that individuals, especially multilingual speakers, have varied skill/comfort levels in different dimensions of language, such as syntax, vocabulary, and phonetics. Additional variation may arise across these dimensions depending on the method of measurement (e.g., picture naming tasks vs. measures of online processing; Birdsong, 2014). Moreover, both monolingual and multilingual individuals experience changes in use and proficiency across the lifespan, complicating the practice of assessing 'nativeness' while relying on a static measure in a single domain.

Leela (**Box 3**), for example, would likely perform "natively" in a Malayalam and English picture naming task, regardless of the comparison group. However, depending on the norms assumed by the researchers, she might not perform "natively" in a phonetic task in English, or a reading comprehension task in Malayalam, because of the contexts in which she learned

TABLE 2 | Questions which can be used to probe various factors of language experience in survey design.

Factor of Interest	Categorical Questions (Useful for Recruitment)	Open-Ended/Gradient Questions (Useful for Survey Design/ Capturing Continuous or Qualitative Differences)
Age and Order of Acquisition	● Did you start learning [Language X] before [Age Y] ● Is [Language X] (one of) your first language(s)?	● At what age did you begin learning [Language X]? *Consider probing* ● How long was [Language X] used before exposure to another language?
Context of Acquisition	● Did you grow up speaking [Language X] in [Region Z]? ● Did you grow up speaking [Language X] at home/in school? ● Were you exposed to [Language X] at home/in school? ● Was [Language X] the language of instruction in school?	● Please list the locations you have lived in, the ages you lived there and the languages you spoke during that time. ● What percentage of the time did you speak/hear [Language X] with your family growing up? ● How many years were you exposed to [Language X] at home/in school? ● How many years was [Language X] the language of instruction at school? *Consider probing* ● Presence of and/or interaction with [Language X]-speaking community networks ● Interest in [Language X] media
Language Proficiency/Usage Practices	● Can you speak/read [Language X]? / Are you a (fluent) [Language X] speaker? / Are you comfortable speaking [Language X]? ● Do you mainly use [Language X] (at home/at work/in daily life)?	● Please rate on a scale of 1–7 your proficiency/fluency/comfort in [speaking/understanding/reading/writing; Language X]. ● What percentage of the time do you speak/hear [Language X] at home/work/school/with friends? *Consider probing* ● Language use in different spheres (e.g., home, work, and school) ● Literacy ● Which language varieties they use
Language Identity/Allegiance	● Are you a [Language X] speaker? ● Do you consider yourself a [Language X] speaker? ● Do you consider [Language X] to be your? [contextually relevant term, e.g., native language, Mother Tongue]?	Please rate on a scale of 1–7 how much you agree with the following statements: ● "[Language X] is my [native language/Mother Tongue/etc.]" ● "[Language X] is my preferred language" ● "I feel a strong connection to [Language X]" *Consider probing* ● Feelings or perceptions of pride, value, community, and nationality related to using or learning [Language X]

both of those languages. In addition, Leela now uses solely English at work; when she was a child, she was almost never in English-only contexts. As such, any assessment of her comfort in using English would likely change significantly across her lifespan. Crucially, despite spending her whole life in a Malayalam-speaking region and having an education background that is quite common for people her age, she would show quite different performance across domains. This not only demonstrates the problems with prioritizing certain measures over others, but also could potentially lead to harm by reinforcing deficit perspectives toward multilingualism.

Potential histories of oppression and marginalization should also be accounted for. Surrain and Luk (2019) note that many standardized measures of language proficiency did not consider multilingual children in their norming process and that many of the standardized normed tests were developed to evaluate for atypical language behavior rather than proficiency. Care must be taken in the use and interpretation of standardized measures of language proficiency to avoid framing differences as deficits. Following Surrain and Luk, who found that only 38% of studies on children and 17% of studies on adults addressed sociolinguistic context in discussions of participants and results, we suggest that these details be reported in publications.

By being explicit about the aspects of language experience that we are concerned with, what assessments we use, and what assumptions our assessments carry, we can help avoid inconsistency. In the same vein, when selecting measures to characterize participants or while performing cross-study comparisons, we suggest researchers pay close attention to how 'nativeness' is defined and measured. Keep in mind that the fact that two studies use the term 'native speaker' does not guarantee that both studies share the same operationalization. The assessments chosen and how they are reformulated to

be more targeted and specific will vary based on the relevant aspects of language experience that the researcher is interested in.

Complicating NATIVENESS in Data Processing and Analysis

Continuous Variables and Linear Regression

Many scholars have noted issues surrounding entrenched "categorical thinking" in psychology and psycholinguistics, marked by sampling few values along a continuum, discretizing continuous measures, and an overreliance on factorial design and ANOVA (for more detailed discussion, see MacCallum et al., 2002; Baayen, 2004; Balota et al., 2004; Young, 2016). In some cases, categorizing continuous variables is purposeful or necessary and reflects a certain research goal; however, problems arise when this approach becomes entrenched and hinders theoretical progress. While dichotomizing continuous predictors generally results in a loss of power, Baayen (2004) notes that in the case where there is a single relevant predictor X, building a factorial contrast for extreme values of X (i.e., assigning the first 5% of ranked values to a "low" condition, and the last 5% to a "high" condition) can increase statistical power. However, this strategy comes at the price of being limited to making generalizations about the extreme ranges of X. A contrast between the high and low groups can be established, but no prediction is possible for the values of X in the intermediate range. Furthermore, as Balota et al. (2004) note, information about the amount of unique variance that a given factor accounts for in a given design should be taken into consideration in our theory formation and selection. Thus, Baayen (2004) recommends that "factorization is useful when obtaining data is costly and when documenting the existence of an effect is the sole purpose of the experiment" (6).

Scholars studying bilingual populations have argued for language proficiency, usage, and acquisition-related factors to be measured and analyzed gradiently. Luk and Bialystok (2013), for example, assert that "bilingualism is not a categorical variable," finding that language proficiency and bilingual usage are two continuous factors which should be included in models characterizing bilinguals (see also Gullifer et al., 2021). Similarly, Ortega (2020), in a paper on heritage language development from a social justice perspective, argues that, in order to properly characterize the language of "heritage" language speakers (cf. Benmamoun et al., 2013), one must take a gradient approach to bilingualism. In addition, some sign language linguists have argued for the use of a continuous measure of nativeness, as opposed to a binary categorical definition (Costello et al., 2008). Treating a predictor, like "language dominance" as a continuous variable in analysis, allows for a more fine-grained analysis which answers questions about whether the more dominant a bilingual is in a given language, the more likely they are to demonstrate certain (psycho)linguistic behaviors (see "Issues with NATIVE SPEAKER when constructing comparison groups and conducting analyses" and Solís-Barroso and Stefanich (2019) for further discussion and examples of continuous vs. categorical language dominance measures). The interpretations offered by continuous analysis may also be easier to align conceptually with the knowledge that certain linguistic measures, like language dominance, are dynamic within the individual.

Individual Variation and Mixed-Effects Modeling

As discussed, there is a lot of heterogeneity in language experiences, and instead of attempting to make (artificially) homogenous groups, an alternative is to conduct individual differences analyses. Some approaches to language take such individual differences as the norm, even in relatively homogenous populations. For example, Dąbrowska (2013) argues from a usage-based perspective that what ends up looking like language-wide grammatical constraints are likely based in the cognitive biases of a subset of individuals, which then get amplified through patterns of transmission and use. Individual differences analyses often focus on factors such as working memory/attention, print exposure, or categorization gradience as predictors of linguistic behavior, predicting long-distance dependency resolution (e.g., Nicenboim et al., 2015), pronoun comprehension (Langlois and Arnold, 2020), or processing of phonetic cues (Ou et al., 2021), respectively. These types of analyses can also be applied to investigate the role of language exposure/use on linguistic knowledge/behavior.

Mixed-effects models, in which random effects allow for sub-group differences (as well as stimulus-derived variation), have become the norm in psycholinguistics. In particular, random by-participant intercepts and slopes account for variation at the individual level, which is typical in (psycho)linguistic data (see Barr et al., 2013). However, beyond simply "factoring out" individual-level variation that is not of interest to the research question, it is also often informative—and potentially crucial—to analyze individual response patterns in addition to group patterns. To illustrate, Tanner et al. (2013) found that while electrophysiological responses at the group level showed a biphasic pattern, no single individual showed that pattern but rather either an N400 or a P600 effect. In line with the individual differences approach, the authors argue that this demonstrates how new insight can be gained "when the cross-subject variability is treated as a source of evidence rather than a source of noise" (Tanner et al., 2013). One method to analyze individual variation is to use random effect coefficients output by mixed-effects model as the response variable, an approach that is increasingly common, for example, in individual-level correlation analyses of speech production and perception patterns (e.g., Pinget et al., 2020; Voeten, 2020). Overall, this approach has the advantage of moving away from categorical thinking and moving toward understanding underlying factors and mechanisms in language processing, which is abound with meaningful variability (e.g., see Yu and Zellou, 2019 for an individual differences approach to phonological processing).

Multivariate Data and Dimensionality Reduction

One way to handle large amounts of detailed demographic and language experience data collected from a questionnaire is to use dimensionality reduction techniques, like factor analysis or principal components analysis, to distill the data from many continuous measures into a smaller, more manageable number

of relevant, orthogonal continuous factors. These factors can then be used as predictors in a regression model. Especially, when dealing with a sociolinguistic context which might be unfamiliar or understudied, allowing the relevant predictors to be inferred from the data in a principled manner, as opposed to data fishing, might be a desirable approach. This allows researchers to model language experience factors without defaulting to researcher-imposed categories; it could be that categories emerge from the data, but this allows that information to be inferred rather than imposed.

For example, Luk and Bialystok (2013) examined the responses of a highly heterogeneous group of 110 bilingual individuals (defined as individuals who had experience using two languages on a daily basis, with English being the dominant language of the community) to an English proficiency and self-report questionnaire, the Language and Social Background Questionnaire (LSBQ). Participants also completed the Peabody Picture Vocabulary Task-III (Dunn and Dunn, 1997) and the Expressive Vocabulary Task (Williams, 1997). Taking the many measures they collected, they conducted a factor analysis which found that language proficiency and bilingual usage are two continuous factors which should be included in models characterizing bilinguals (at least in a context where there is a dominant language in the society at large). For details of their analysis, we direct the reader to their paper, but this provides an example of how such analyses can be used to address theoretical questions about a heterogeneous group of participants. Other related analyses, such as principal components analysis, have long been used in studies of typology and variation (e.g., Abdi and Williams, 2010).

Emergent Groups and Clustering

An alternative approach to multivariate data, rather than identifying latent variables, is to use clustering techniques to classify individuals and identify emergent groups. According to Garcia-Dias et al. (2020), clustering is "a type of unsupervised learning which aims to find the most natural way of grouping a dataset" based on similarity across various dimensions. Clusters can then be interpreted by the researchers based on the contributing variables, and external validation can further be conducted to assess the extent to which each emergent cluster aligns with independently known variables. In other words, researchers can use clustering as a data-driven approach to identifying groups, if they exist, based on language experience or behavior rather than rely on predetermined and/or dichotomized categories based on 'nativeness' or other constructs.

Clustering can be used to identify "natural" groups (in a more informed manner than techniques such as median split) which are applied to test hypotheses or make comparisons. In some cases, groups can be predicted. For example, Novick et al. (2014) used clustering to identify "responders" and "non-responders" to an *n*-back task to investigate how responsiveness to cognitive control training influences recovery from misanalysis of sentence structure (i.e., garden-path sentences). In other cases, groups can be fully emergent. Chiarello et al. (2012), for instance, used clustering to identify four distinct subgroups of college-aged readers based in part on reading skill and then investigated the neurological correlates. Clustering is also generally useful for exploratory aspects of the research process, such as identifying individuals with similar response patterns to aid in data interpretation (e.g., sentence comprehension ability in aphasiac individuals or mono- and multilingual children, Caplan et al., 1985; Filippi et al., 2020). In addition, external validation can be a way of both exploring the data and testing hypotheses, often *via* the lens of individual differences (see "Individual Variation and Mixed-Effects Modeling"). In a study of second dialect acquisition, Voeten (2020) provides an example of using residential history as an external variable to examine the extent to which migrants who moved from Belgium to the Netherlands had adopted more Netherlandic-like vowels or retained more Flemish-like vowels (i.e., whether migrants' vowel production measures clustered with those externally identified to be raised in the Netherlands or Flanders).

ACTIONABLE RECOMMENDATIONS

Like Section "Assessment and Implications of Current Practices", our actionable recommendations are divided into three sections that correspond to different stages of conducting research: conceptualization of research questions, characterization of language experience in experimental materials, and analysis of data. We begin with suggestions on how to conceptualize research questions and encourage a pause for reflection about the underlying assumptions of experimental constructs. The remainder of the section proposes ways to move beyond NATIVE SPEAKER with regard to various aspects of an experiment, including participant recruitment, stimuli development, task design, and assessment or screening question selection as well as during data processing and analysis. Throughout, we offer concrete actions to take along with illustrative examples to help readers apply these recommendations to their own research. However, not all laboratories or researchers have access to the same resources, whether that be physical space, time, funding, personnel, populations of interest, etc., and may not be able to follow all of these recommendations. We encourage researchers to do the best they can give their constraints. Researchers should prioritize the recommendations that best suit their research questions and experiments (see "Alternatives to NATIVE SPEAKER in Conceptualization" for help identifying which aspects of NATIVE SPEAKER are important for one's research).

Alternatives to NATIVE SPEAKER in Conceptualization

Before designing an experiment, researchers need to take time to think critically about both their own assumptions about NATIVE SPEAKERS and the questions they want to investigate. Understanding one's own biases before engaging in research will result in more ethical and reproducible science. Throughout this questioning process, we encourage researchers to keep a log of the questions they ask and their answers. This log can

help when writing up one's theoretical viewpoint in a paper, explaining why or why not one included a variable in an experiment, as well as to see how one's viewpoint has changed over time. See **Figure 1** for a consolidated list of suggested questions.

Please note that we are not encouraging researchers to strictly define NATIVE SPEAKER and continue using the term, but instead to think about the more granulated aspects of language experience that they use NATIVE SPEAKER to mean. This is an iterative process, and researchers will benefit from engaging in this process throughout their career.

Reflection

First and foremost, we recommend that researchers ask themselves what their assumptions about NATIVE SPEAKERS are. This will help pinpoint the aspect(s) of language experience they are interested in and which they think are relevant for their overall research. To do this, researchers could start by concretely defining NATIVE SPEAKER in their own terms. For each of the attributes that the researcher lists, they should ask what the attribute implies. For example, if a researcher lists that a NATIVE SPEAKER "learned language X before age Y," then this attribute implies a critical period. The researcher should evaluate whether or not they agree with this implication, determine if it is an important aspect of language experience, and update their definition accordingly.

Researchers should then evaluate the theoretical beliefs that they hold about NATIVE SPEAKERS. One should ask themselves what theories they do or do not support and list the assumptions that those theories make about language experience. Then, the researcher can determine if they agree with those assumptions. The researcher may notice a trend among the assumptions that they agree with. For example, a researcher may support theories that imply that someone is a NATIVE SPEAKER of X if X was the language of instruction at school. This indicates that the researcher views language of instruction as an important aspect of language experience. Crucially, a researcher should then ask "Would including a participant with a different language profile change my theoretical predictions?" If the answer is "Yes," then the researcher should consider what aspect of a participant's language profile would have to be different (and how different) for the researcher to see a change in their predictions. Once a researcher has completed this line of questioning, they will be more equipped to determine the aspects of language experience they consider to be important for their research as a whole. Identifying one's underlying assumptions can then inform recruitment, experiment design, and data analysis.

During this reflection, some may feel resistance to shifting away from NATIVE SPEAKER, which may have been a functional or central concept in their past research. We do not deny that certain commonalities may be observed among those broadly considered NATIVE SPEAKERS (or L1 speakers) by researchers. However, these communalities may be more precisely captured by factors of language history, proficiency, and/or identity, meaning that a move away from fuzzy categories is a move toward clarity of mechanisms. Given this context, we ask that researchers consider carefully whether the implicit and explicit exclusion of marginalized and minoritized populations through the usage of this term is necessary or justified for their particular research program or questions. It is crucial for researchers to be explicit about the types of language experience which are important to their research questions in order to avoid reproducing normative assumptions about who gets to be a NATIVE SPEAKER. By pulling away from the term NATIVE SPEAKER as a proxy variable, researchers can reduce harm and begin to better align their research and theory with non-normative contexts of language learning and use.

Research Design

After researchers have analyzed their own beliefs, they can begin designing experiments that take into account their assumptions. Researchers often begin the research conceptualization process with a general question that they translate into a specific, operationalized hypothesis. If researchers have gone through the initial reflection process, then they already know what aspects of language experience are relevant for their research, and what aspects are important for specific questions. However, it is important to ask follow-up questions for each new experiment. Researchers should engage in this process for each new experiment that they design.

The first question a researcher should ask is as follows: What aspect of language experience is important for this specific research question? If they are not able to identify an aspect, they should go through the self-exploration process in Section "Reflection" again. Once researchers have identified what aspect of language experience is important for their question, they should identify why it is important. This question has two goals: (1) It situates the work in a theoretical context, and (2) it ensures that the construct is relevant. When researchers tie their experiments to broader theories, they should again evaluate what these theories imply about NATIVE SPEAKERS and whether or not they agree. If researchers are not able to identify why a construct is important, then they should consider using a different aspect of language experience (though note that this may not be relevant for exploratory research, where there may not be strong evidence for the importance of a construct). At this stage, it is also beneficial to determine what predictions come from a researcher's chosen aspect of language experience. While this should be a standard part of experiment design, thinking through possible results is particularly beneficial in the case of research derived from the concept of NATIVE SPEAKER because it can indicate whether existing constructs are informative and relevant.

One should also consider whether including speakers with different language profiles would affect the data, and if so, how. Similar to whether a different language profile would impact theoretical predictions, this question evaluates both the scope of the empirical question and how susceptible the experiment design is to heterogeneity within speaker groups. If slight variations in speaker profile change the predictions, then the natural heterogeneity of participant groups will lead to differences in results. Exploring how one's data would change by including participants with different language profiles may

reveal relevant aspects of language experience the researcher had not previously considered.

Considering how language profiles affect data leads nicely to the next important question: Who is the researcher *excluding* and why? If researchers are excluding a group due to a specific aspect of language, we encourage researchers to tailor their screening questions to *measure* that aspect of language instead of blanket-eliminating a widely heterogeneous group of subjects (see "Alternatives to NATIVE SPEAKER in data processing and analysis" for more information). Researchers should consider what their data and predictions would look like if they did include these speaker groups and contexts.

Lastly, researchers should dedicate a considerable amount of time to evaluate how they structure their research questions and situate them in a theoretical framework. As discussed previously, many existing theories make assumptions about NATIVE SPEAKER and inherently imply other debated frameworks. By carefully examining the assumptions underlying theory, researchers can avoid vagueness and enhance replicability. Assumptions about NATIVE SPEAKER as a concept may also exclude certain participants from research, perpetuating harmful stereotypes and deficit models. We call on researchers to establish their research in explicit theories of language learning, exposure, and usage. Within this solid framework, researchers can then clearly state what they are testing and how it relates to the variables within the theoretical framework. The participants' explicit, specific, and measurable language experience should be the driving variables in theory development and experiment design.

Alternatives to NATIVE SPEAKER in Recruitment, Tasks, and Surveys

Various aspects of an experiment may use the term NATIVE SPEAKER or assume the construct as given. In this section, we provide best practices to move beyond this when designing materials for participant recruitment, experimental tasks, and surveys, as well as when reporting methodological details in a publication.

Participants and Groups

First and foremost, we recommend that researchers actively avoid describing participants as '(non-)native speakers' in any part of the experiment, including the use of this concept as a criterion for assigning participants to target and comparison groups. Instead, the best practice is for researchers to, based on their specific research question, specify targeted aspects of linguistic experience in detail (for recommended reflection questions, see "Alternatives to NATIVE SPEAKER in Conceptualization"). These concrete characteristics can then be used in recruitment and analysis, promoting clarity and transparency.

For example, when asking about the effect of language experience on acceptability of resumptive pronouns in Egyptian Arabic, one might construct comparison groups differently based on one's particular assumptions about the role of language experience. If a researcher thinks that speaking Egyptian Arabic as the primary language is what matters, that would be one way of recruiting and characterizing a comparison group. Alternatively, if a researcher thinks age of acquisition is what matters, recruitment could take place based on whether participants had grown up speaking Egyptian Arabic. Crucially,

Personal reflection:

1. **In my opinion, who is a NATIVE SPEAKER?**
   a. What do these attributes imply?
   b. Do I agree with these implications?
2. **What theories regarding language experience do I support/not support?**
   a. What do these theories say (explicitly and implicitly) about NATIVE SPEAKERS?
   b. Do I agree with these conclusions?
   c. Would including a speaker with a different profile change my predictions?
      i. If yes, what aspect of the language profile would have to be different, and how different would it have to be?
      ii. How might I need to change the scope of my claim given the profiles of individuals my work is targeting?

End Goal: What aspect of language experience do I consider important for my research as a whole?

Research design reflection:

1. **What aspect of language experience is important for this research question?**
   a. Why is this aspect important?
   b. What predictions (if any) does it make?
   c. Would a speaker with language profile Y affect my data? How?
2. **Who is excluded from my research?**
   a. What acquisition and use contexts am I including and excluding?
   b. What would happen if I included them?
3. **What assumptions and theories does my design explicitly and implicitly espouse?**
   a. Are differences presented as deficits? Could they be interpreted as such? Do I actively work against deficit perspectives, especially towards minoritized communities/individuals?

End Goal: What aspect of language experience is important for this experiment and what does my design imply?

FIGURE 1 | Reflection and conceptualization questions to ask during the research process. Please note that we are not encouraging researchers to define NATIVE SPEAKER in order to aid in continuing to use the term, but instead to think about the more granulated aspects of language experience that they use NATIVE SPEAKER to mean. This is an iterative process, and researchers will benefit from engaging in this process throughout the research process and their careers.

these different hypotheses would result in potentially different groups of participants, and asking about 'native speaker' status would result in a third, less-specified group.

Some of this information can be collected after the experiment is completed, in a post-experiment questionnaire, but, when possible, conducting norming studies (finding out what types of language experience are common in the local context), looking to census data, or making hypotheses based on ethnographic or sociolinguistic research is also a good practice (cf., Chung et al., 2012; for recommended questions to consider, see "Surveys and Recruitment").

Stimuli and Norming

Likewise, researchers should make sure to stipulate beforehand, and collect detailed information about, the demographic and language experience characteristics of individuals recruited to record auditory stimuli or to participate in norming/coding of experimental items. The experimenters can then use this information in the interpretation and discussion of results. Ensuring that this information is also reported in publications serves to transparently communicate the context for readers to potentially replicate the study materials.

Crucially, we are not asking that researchers necessarily report any and all information they might have (though reporting such information in supplementary materials or appendices is an option); rather, we ask that researchers aim to have clear, *a priori* expectations about which aspects of language experience are relevant for their particular research context (see "Surveys and Recruitment" for some ideas). In the case of recording auditory stimuli, some relevant elements to report are (i) whether the person is from the same community as the participants, (ii) how their language experience might match or deviate from that of the participants, and (iii) when possible, any information about whether and how that might be perceived by the participants.

For example, if a researcher is conducting a study on constituent ordering preferences in Hindi-Urdu speakers living in the United States, rather than reporting that 'a native speaker of Hindi-Urdu was recorded for auditory stimuli,' a better practice would be to report details, including the location and linguistic context in which the speaker grew up and currently resides, the speaker's regional/cultural variety of Hindi-Urdu, and their language usage practices, including potential multilingualism. Similarly, for a researcher conducting a study on processing of 'foreign-accented' speech, rather than reporting that 'ten non-native speakers of English recorded sentences,' it would be better practice to report details (in the main text or in a table) about each speaker's regional background, linguistic history, and perceived degree of accentedness.

Tasks and Assessments

When choosing assessments to characterize participants, researchers should keep the following question in mind: Is this assessment measuring 'nativeness' as I operationalize the term, or is this assessment capturing others' ideologies about what a 'native speaker' is? When we ask participants to self-report their "native" status, in reality, the results capture their ideologies of what a 'native speaker' is and whether they fit into that definition. The same reflection question must be used when considering apparently objective measures. Take, for example, 'nativeness' accent ratings. In this type of assessment, speech samples from the participants are given to raters (usually members of the target speech community) who then judge how "native" a speaker sounds. While accent ratings appear to be collecting objective data, in reality, this measure too is capturing the rater's subjective ideologies about what a 'native speaker' should sound like and whether their perception of the speaker matches their ideologies of nativeness.'

To avoid assessing ideologies about NATIVENESS and leaving room for interpretation of what this term means, we recommend that our assumptions as researchers about NATIVENESS be made explicit in the assessments we select. If, for example, a researcher were to be interested in determining whether participants grew up speaking the "native" language at home and at school during early childhood, rather than asking an ambiguous question, such as "Are you a native speaker of German?" one could ask more targeted and explicit questions, such as "Did you grow up only speaking German and spoke it at home and school?" (see **Table 2** for more examples). The same reformulation can be made to assessments like the 'native accent' ratings. For example, instead of asking "Does this person sound like a native speaker of Spanish?" a more specific question could be "Does this person sound like they grew up in Mexico?", "Does this person sound like a local?", or "Does this person sound like they are a monolingual speaker of Spanish?". The same caveats about limiting participant pools apply here; we are not advocating that researchers stick to ostensibly monolingual participants. Rather, this is a demonstration of less harmful and more accurate ways of asking questions about language experience.

As for assessments of proficiency, we have given examples throughout this paper of how uncritical use of proficiency measures can be damaging and misleading. However, if a notion of proficiency is crucial to the research question, we advise that researchers keep in mind that proficiency can vary across domains (e.g., speaking, understanding, reading, and writing), and we urge that proficiency be properly contextualized, taking language access and structures of oppression into account. Much like Ortega (2020) suggests in her social justice-focused review of heritage language development, psycholinguists can guard against assumptions of proficiency that privilege hegemonic monolingual norms by ensuring their research considers the myriad ways proficiency develops and is demonstrated across a person's life course. Considering how local language ideologies and institutional opportunities and impediments influence participants' language use in different domains will lead to a more nuanced picture of proficiency.

Surveys and Recruitment

The questions in **Table 2** give examples of how researchers might question and characterize aspects of language experience relevant to their research purpose without relying on implicit assumptions about 'native speakerhood.' Consider that while focusing on a small number of categorical distinctions may be useful for recruiting participants in a straightforward and

practical way, it may be more helpful for theorization and analysis to capture these variables with more gradient measures, as discussed in Section "Complicating NATIVENESS in Data Processing and Analysis".

Whether or not one's study design involves explicit predictions or analyses based on participants' language experience, the sociolinguistic context should inform how researchers choose questions for recruiting and characterizing participants and how they interpret and report responses. Norms of multilingualism and schooling in the local context may affect to what degree a given question about language experience will generate a homogeneous sample. The desirability of homogeneity will depend on the purposes of the study, but be cautious of the trade-off between the homogeneity of a given sample and the generalizability of an observed effect. We caution against making broad claims about language organization and behavior based on studies drawing on samples from relatively homogenous (and WEIRD) speech communities. However, precisely, reporting the language and cultural background of one's participants (as discussed in "Participants and Groups" and "Stimuli and Norming") can help inform directions for expansion in the future research.

Here, we give some suggestions for the types of information which might be relevant to take into account when contextualizing results, and therefore the types of information which researchers can ask about and report in publications. For example, is the language or variety stigmatized, either locally or by the larger society? Is the community minoritized (even if the language/variety is not stigmatized)? Are the speakers (and perhaps by extension particular linguistic features) racialized? What is the incidence of multilingualism in the community? How does schooling look typically, and what are the associated language policies? What counts as being multilingual for the local context? Are the language boundaries which linguists can perceive relevant for the speakers themselves (cf. Otheguy et al., 2015)? Looking to the boxes in this paper can give some examples of how speakers and communities might vary, and the types of information which would be relevant for understanding how speakers may or may not be typical of the populations of interest.

Alternatives to NATIVE SPEAKER in Data Processing and Analysis

Rather than comparing groups of participants based on NATIVE SPEAKER status, we recommend that researchers compare response measures based on more transparent and targeted variables of language experience relevant to their research question. As part of this move away from categorical analyses, we encourage researchers to take an individual differences perspective to foreground individual-level patterns in their data. There are several ways this could be implemented. In a context where researchers recruit from the same pool of participants, expecting heterogeneity of experience but not expecting qualitatively separable groups, one could examine data for apparent outliers and attempt to interpret these based on language experience factors. Alternatively, one could examine whether the dependent variables are multimodally distributed, and conduct *post-hoc* analyses connecting those patterns with information about language experience. Finally, even in contexts where there might be qualitatively separable groups, researchers can examine how relevant language experience factors might contribute to how individuals do or do not map on to group-level patterns.

We briefly direct the reader to methods of analysis which align with the theoretical moves we are advocating to accurately and fairly represent our research populations, namely, using continuous variables instead of categories in linear regression and taking into account individual differences using mixed-effects modeling (Regression Analysis), as well as using dimensionality reduction techniques on many continuous variables and using clustering techniques to split the data into observed rather than predetermined groups (Multivariate Data). We recognize that these statistical tools may not be available or applicable to all researchers, particularly those who work with smaller sample sizes (e.g., due to constraints of the participant population); in such cases, it might be necessary to consider whether non-parametric statistical methods or qualitative analyses are more appropriate.[9] Our recommendation is that researchers thoughtfully consider using the alternative methods and approaches that are relevant for their particular research context, given their practical limitations.

Regression Analysis

Regression analysis is widely used in psycholinguistics, and it allows us to account for continuous predictors. To take a concrete example, let us say Dr. A is interested in differences in lexical processing based on language experience (determined through the researcher's own reflection during conceptualization; see "Alternatives to NATIVE SPEAKER in Conceptualization"). Dr. A may recruit participants who are all currently residing in Germany and have either German or English as their L1. Along with a German lexical decision task, they ask their participants about their age of acquisition and self-rated language usage and proficiency in a post-experimental questionnaire. One or more of these variables can be added as continuous predictors into a regression model.

Familiarity with regression-based approaches to analysis is necessary when dealing with continuous predictors, and unfamiliarity and lack of training with regression-based statistics may explain some of the continued "categorical thinking" in the field. We recommend that researchers follow the current norm for modeling psycholinguistic data, which is to use mixed-effects regression models (also called random-effects, hierarchical, and multilevel models; Gelman and Hill, 2006, p. 2) to control for participant heterogeneity (i.e., individual

[9]Due to disciplinary norms and training, psycholinguists may not be used to thinking about qualitative analyses, including descriptive statistics or visual interpretation (as relevant), as an option. We are not suggesting that qualitative analyses are inferior or a "last resort," but rather they yield different types of information. We hope psycholinguists can consider this an additional tool in their analytic toolkit. As discussed in the conclusion, we recognize this would require structural changes beyond the control of an individual, but being able to include smaller populations without obscuring meaningful variation would be a great advantage.

variability). To continue the example above, Dr. A should, at the very least, include by-participant random intercepts, which allow individual participants to be modeled with different mean response times, as well as by-participant random slopes, which allow for individually variable patterns of response time to experimental conditions (e.g., real vs. nonce words). Some resources for learning and using regression analyses include Winter (2019) *Statistics for Linguists: An Introduction Using R*, Baayen (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*, and Gelman and Hill (2006) *Data Analysis Using Regression and Multilevel/Hierarchical Models* (see also Barr et al., 2013 and Bates et al., 2018 on conventions for constructing mixed-effects models).

Multivariate Data

Dimensionality reduction and clustering techniques allow one to identify patterns between several variables simultaneously, which can be particularly helpful when handling and making sense of large amounts of language experience data. When multiple predictors are considered in regression modeling, correlations between predictors (i.e., collinearity) can cause adverse consequences for analysis. This is something to keep in mind when dealing with many language experience variables, because we may expect several variables of language experience to correlate with each other, in addition to being correlated with the response measure. For example, age of acquisition is often correlated with measures of proficiency (Birdsong, 2005). So, Dr. A may choose to select only one of the highly correlated variables in their data set to enter as a predictor in the model (especially when these variables and their consequences are not of primary interest), or they may choose to perform more sophisticated multivariate statistical techniques to work around this problem. Tomaschek et al. (2018) address several strategies for diagnosing and addressing collinearity in multivariate linguistic data, of which dimensionality reduction is one solution.

Dimensionality reduction techniques include factor analysis (FA) and principal components analysis (PCA). These techniques are similar, but different in their approaches and assumptions. PCA is usually preferred when the goal is simply to reduce correlated observed variables to a smaller set of composite variables, whereas FA is preferred when the goal is to detect underlying factors influencing the responses on the observed variables (for a more detailed but still approachable comparison of these methods, see Brown, 2009). Luk and Bialystok (2013) provide their reasoning for using FA instead of PCA in precisely these terms; they were interested in an assumed underlying causal relationship between the observed variables and the latent factors of interest, which were related to aspects of bilingual experience. Hypothetically, Dr. A may choose to do a factor analysis and find two latent factors which they interpret as roughly corresponding to language history (age of acquisition, proficiency, and past usage) and current language usage. For an introduction to factor analysis, see Thompson (2004) and Formann (2014) for latent class analysis. For an introduction to FA and PCA (among other multivariate analysis techniques) in R, see Chapter 5 of Baayen (2008).

Whereas dimensionality reduction aims to identify and reduce irrelevant or redundant variables, clustering aims to identify natural groupings of data points based on similarity. Language experience variables (e.g., from a screening task or language experience questionnaire responses) may not always be evenly distributed across a continuous range. Data may instead be multimodal such that certain participants are particularly similar to each other, representing groups who are qualitatively different, at least in that particular context or participant sample. In this case, participant groups could be identified in a data-driven way *via* cluster analysis, interpreted relative to the language experience profile per cluster, and then used as a categorical variable in planned analyses. To illustrate, if Dr. A performed a cluster analysis on each participant based on their reported German age of acquisition, usage rating and proficiency rating, this may hypothetically result in three clusters with distinct language experience profiles: early learners with high usage and high proficiency, later learners with high usage and high proficiency, and later learners with low usage but high proficiency.

Before clustering, diagnostics should be used to confirm that the data indeed are multimodal. There are various types of cluster methods, including the broad classes of hierarchical methods, where each individual begins as its own cluster and is grouped with progressively more clusters (e.g., Ward's method) and partitioning methods, where number of clusters are prespecified and individuals are assigned to a particular cluster (e.g., K-means). For more information about cluster analysis, Jain (2010) provides an overview of the method (see also Chapter 5 of Baayen, 2008). Other resources include Garcia-Dias et al. (2020), which focuses on K-means clustering, and Clatworthy et al. (2005), which reports on how clustering analyses have been used in health psychology.

CONCLUSION

It is clear that the concept of NATIVE SPEAKER can be harmful to both psycholinguistic research and the populations we study. The term NATIVE SPEAKER has been problematized in different fields, and some solutions have been to repurpose the term. However, these approaches simply redefine NATIVE SPEAKER to include or exclude certain populations (Costello et al., 2008; Benmamoun et al., 2013; Rothman and Treffers-Daller, 2014) and many of the same problems remain. We encourage researchers to abandon the term altogether and join our colleagues in adjacent fields to adopt a more nuanced view of language experience. We anticipate that the recommendations in Section "Actionable Recommendations" can improve research at all stages: theory construction, experiment design, participant recruitment, stimuli creation, and data analysis. To reiterate, researchers should explicitly define which aspects of language experience they are investigating, recruit individuals and select assessments that are targeted to these aspects, ask specific questions to understand participants' language experience, account for the sociolinguistic context in which the questions are asked and how they are

interpreted, and use gradient measures instead of categorical variables when relevant.

As we reflect on these recommendations, recall the point made in Section "Actionable Recommendations"—as researchers may not be able to practically implement all of these all at once, they should prioritize the changes which are most relevant for their research questions and the contexts in which they work. However, individuals and individual laboratories cannot on their own be responsible for shifting away from this concept and the associated baggage. We recognize there are structural barriers to implementing some of these recommendations, such as expectations of reviewers, access to funding to collect requisite data, and necessarily small sample sizes when working with certain populations. Of course, we psycholinguists are part of these structures to different degrees and can work within our spheres of influence to make a difference. This could look like advocating for less harmful ways of characterizing and recruiting participants as part of local/institutional ethics boards, program committees, journal editorial boards, and in classroom or other training contexts. Structural changes are necessary and must accompany the individual-level changes which are the focus of this paper. We hope these recommendations will not only improve research design and analysis, but also contribute to the co-creation of a more just and inclusive field.

AUTHOR CONTRIBUTIONS

All authors contributed to the conceptualization and writing of this paper. The most substantive contributions of each individual are listed as follows. LC: Vagueness, Issues With NATIVE SPEAKER at the Stages of Conceptualization and Design, Issues With NATIVE SPEAKER When Constructing Comparison Groups and Conducting Analyses, Complicating NATIVENESS in Conceptualization, Continuous Variables and Linear Regression, Multivariate Data and Dimensionality Reduction, Emergent Groups and Clustering, Participants and Groups, Stimuli and Norming, Tasks, and Assessments, Regression Analysis, and Multivariate Data. DB: Issues With NATIVE SPEAKER When Constructing Comparison Groups and Conducting Analyses, Continuous Variables and Linear Regression, Individual Variation and Mixed-Effects Modeling, Multivariate Data and Dimensionality Reduction, Emergent Groups and Clustering, Surveys and Recruitment, preparation of **Table 2**, Regression Analysis, and Multivariate Data. NV: Continuous Variables and Linear Regression, Complicating NATIVENESS in Conceptualization, Reflection, Research Design, preparation of **Figure 1**, and Conclusion. CS-B: Vagueness, preparation of **Table 1**, Complicating NATIVENESS in Recruitment, Tasks, and Surveys, and Tasks and Assessments. AM: Harm, Issues With NATIVE SPEAKER at the Stages of Conceptualization and Design, and Complicating NATIVENESS in Recruitment, Tasks, and Surveys. SN: Harm, Issues With NATIVE SPEAKER at the Stages of Conceptualization and Design, Issues With NATIVE SPEAKER When Constructing Comparison Groups and Conducting Analyses, Complicating NATIVENESS in Conceptualization, Complicating NATIVENESS in Recruitment, Tasks, and Surveys, Multivariate Data and Dimensionality Reduction, Participants and Groups, Stimuli and Norming, Stimuli and Norming, and Conclusion.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their invaluable feedback, as well as Anna Lim, the University of Michigan Psychology Department Methods Hour, the University of Michigan Psycholinguistics group, the Speech Production Lab at the University of Oregon, the Language Acquisition and Visual Attention Lab at Boston University, Carl Börstell, Elaine Francis, and the students in LING 740 for their feedback on previous stages of this work.

REFERENCES

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *WIREs Comput. Stat.* 2, 433–459. doi: 10.1002/wics.101

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

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The blank screen paradigm. *Cognition* 93, B79–B87. doi: 10.1016/j.cognition.2004.02.005

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

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers* 1, 1–45.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge University Press.

Babel, M., and Russell, J. (2015). Expectations and speech intelligibility. *J. Acoust. Soc. Am.* 137, 2823–2833. doi: 10.1121/1.4919317

Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust.Soc. Am.* 133, EL174–EL180. doi: 10.1121/1.4789864

Baese-Berk, M. M., and Morrill, T. H. (2015). Speaking rate consistency in native and non-native speakers of English. *J. Acoust.Soc. Am.* 138, EL223–EL228. doi: 10.1121/1.4929622

Balari, S., and Lorenzo, G. (2015). Should it stay or should it go? A critical reflection on the critical period for language. *Biol. Ther. Dent.* 9, 8–42.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. Gen.* 133, 283–316. doi: 10.1037/0096-3445.133.2.283

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2018). Parsimonious Mixed Models (Version 2). *arXiv* [Preprint].

Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Heritage languages and their speakers: opportunities and challenges for linguistics. *Theoretical Ling.* 39, 121–189. doi: 10.1515/tl-2013-0009

Bent, T., Baese-Berk, M., Borrie, S. A., and McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *J. Acoust.Soc. Am.* 140, 3775–3786. doi: 10.1121/1.4966677

Birdsong, D. (2005). Nativelikeness and non-nativelikeness in L2A research. *De Gruyter Mouton* 43, 319–328. doi: 10.1515/iral.2005.43.4.319

Birdsong, D. (2014). Dominance and age in bilingualism. *Appl. Linguis.* 35, 374–392. doi: 10.1093/applin/amu031

Boltokova, D. (2017). "Will the real semi-speaker please stand up?" language vitality, semi-speakers, and problems of enumeration in the Canadian north. *Anthropologica* 59, 12–27. doi: 10.3138/anth.591.T03

Bosker, H. R., Quené, H., Sanders, T., and de Jong, N. H. (2014). The perception of fluency in native and nonnative speech. *Lang. Learn.* 64, 579–614. doi: 10.1111/lang.12067

Briggs, C. L. (1986). *Learning How to Ask: A Sociolinguistic Appraisal of the Role of the Interview in Social Science Research*. Cambridge, UK: Cambridge University Press.

Brown, J. D. (2009). Principal components analysis and exploratory factor analysis – definitions, differences and choices. *Statistics* 13, 26–30.

Bucholtz, M. (2003). Sociolinguistic nostalgia and the authentication of identity. *J. Sociolinguistics.* 7, 398–416. doi: 10.1111/1467-9481.00232

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in *nature* and *science* between 2010 and 2015. *Nat. Hum. Behav.* 2, 637–644. doi: 10.1038/s41562-018-0399-z

Caplan, D., Baker, C., and Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition* 21, 117–175. doi: 10.1016/0010-0277(85)90048-4

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

Chiarello, C., Welcome, S. E., and Leonard, C. M. (2012). Individual differences in reading skill and language lateralisation: A cluster analysis. *Laterality* 17, 225–251. doi: 10.1080/1357650X.2011.561860

Chomsky, N. (1957). *Syntactic Structures*. Berlin, New York: De Gruyter Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

Chung, S., Borja, M. F., and Wagers, M. (2012). "Bridging methodologies: Experimental syntax in the Pacific [Conference presentation]." in *The 86th Annual Meeting of the Linguistic Society of America*, January 7, 2012; Portland, Oregon.

Clancy, K. B., and Davis, J. L. (2019). Soylent is people, and WEIRD is white: biological anthropology, whiteness, and the limits of the WEIRD. *Annu. Rev. Anthropol.* 48, 169–186. doi: 10.1146/annurev-anthro-102218-011133

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., and Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *Br. J. Health Psychol.* 10, 329–358. doi: 10.1348/135910705X25697

Cohen, J. (1983). The cost of dichotomization. *Appl. Psychol. Meas.* 7, 249–253. doi: 10.1177/014662168300700301

Cook, V. (1999). Going Beyond the native speaker in language teaching. *TESOL Q.* 33, 185–209. doi: 10.2307/3587717

Costello, B., Fernández, J., and Landa, A. (2008). "The non- (existent) native signer: Sign language research in a small deaf population." in *Sign Languages: Spinning and Unraveling the Past, Present and Future. TISLR9, Forty-Five Papers and Three Posters From the 9th Theoretical Issues in Sign Language Research Conference*. ed. R. M. de Quadros, 77–94.

Croft, W. (2013). "Radical construction grammar," in *The Oxford Handbook of Construction Grammar*. eds. T. Hoffmann and G. Trousdale (New York, NY: Oxford University Press).

Dąbrowska, E. (2013). Functional constraints, usage, and mental grammars: A study of speakers' intuitions about questions with long-distance dependencies. *Cognitive Ling.* 24, 633–665. doi: 10.1515/cog-2013-0022

Dahl, Ö. (2015). How WEIRD are WALS languages? Conference presentation. Diversity Linguistics: Retrospect and Prospect; Leipzig, Germany.

Debenport, E. (2011). As the Rez turns: anomalies within and beyond the boundaries of a Pueblo community. *Am. Indian Culture Res. J.* 35, 87–110. doi: 10.17953/aicr.35.2.e22v33412156010g

Degraff, M. (2005). Linguists' most dangerous myth: The fallacy of creole exceptionalism. *Lang. Soc.* 34, 533–549. doi: 10.1017/S0047404505050207

Dewaele, J.-M. (2018a). Glimpses of semantic restructuring of English emotion-laden words of American English L1 users residing outside the USA. *Ling. Approaches Bilingu.* 8, 320–342. doi: 10.1075/lab.15046.dew

Dewaele, J.-M. (2018b). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker'. *Appl. Linguis.* 39, 236–240. doi: 10.1093/applin/amw055

Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-III*. Circle Pines, MN: American Guidance Service.

Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–448. doi: 10.1017/S0140525X0999094X

Faez, F. (2011). Are you a native speaker of English? Moving Beyond a simplistic dichotomy. *Crit. Inq. Lang. Stud.* 8, 378–399. doi: 10.1080/15427587.2011.615708

Filippi, R., Ceccolini, A., Periche-Tomas, E., Papageorgiou, A., and Bright, P. (2020). Developmental trajectories of control of verbal and non-verbal interference in speech comprehension in monolingual and multilingual children. *Cognition* 200:104252. doi: 10.1016/j.cognition.2020.104252

Fisher, J. N., Mirus, G., and Napoli, D. J. (2018). "Sticky: taboo topics in deaf communities," in *The Oxford Handbook of Taboo Words and Language*. ed. K. Allan (New York, NY: Oxford University Press), 139–159.

Formann, A. K. (2014). "Latent class analysis," in *Wiley StatsRef: Statistics Reference Online*. eds. N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. Teugels (Hoboken, New Jersey: American Cancer Society).

Garcia-Dias, R., Vieira, S., Lopez Pinaya, W. H., and Mechelli, A. (2020). "Clustering analysis," in *Machine Learning*. eds. A. Mechelli and S. Vieira (Cambridge, Massachusetts: Academic Press), 227–247.

Gelman, A., and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gerald, J. P. B. (2020). Worth the risk: towards Decentring whiteness in English language teaching. *BC TEAL J.* 5, 44–54. doi: 10.14288/bctj.v5i1.345

Green, L. (2004). "Syntactic and Semantic Patterns in Child African American English." in *Proceedings of the Eleventh Annual Symposium About Language and Society—Austin, Texas Linguistic Forum. Vol. 47*. eds. W. F. Chiang, E. Chun, L. Mahalingappa, and S. Mehus, 55–69.

Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A., Giroud, N., Coulter, K., et al. (2021). Bilingual language experience as a multidimensional spectrum: associations with objective and subjective language proficiency. *Appl. Psycholinguist.* 42, 245–278. doi: 10.1017/S0142716420000521

Hall, W. C., Levin, L. L., and Anderson, M. L. (2017). Language deprivation syndrome: A possible neurodevelopmental disorder with sociocultural origins. *Soc. Psychiatry Psychiatr. Epidemiol.* 52, 761–776. doi: 10.1007/s00127-017-1351-7

Han, C., Musolino, J., and Lidz, J. (2016). Endogenous sources of variation in language acquisition. *Proceed. Nat. Acad. Sci.* 113, 942–947. doi: 10.1073/pnas.1517094113

Hanulíková, A., van Alphen, P. M., van Goch, M. M., and Weber, A. (2012). When one Person's mistake is Another's standard usage: The effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Hill, M. (2006). Children's voices on ways of having a voice: Children's and young people's perspectives on methods used in research and consultation. *Childhood* 13, 69–89. doi: 10.1177/0907568206059972

Holliday, A. (2006). Native-speakerism. *ELT J.* 60, 385–387. doi: 10.1093/elt/ccl030

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recogn. Lett.* 31, 651–666. doi: 10.1016/j.patrec.2009.09.011

Johns, C. L., Jahn, A. A., Jones, H. R., Kush, D., Molfese, P. J., Dyke, J. A. V., et al. (2018). Individual differences in decoding skill, print exposure, and cortical structure in young adults. *Lang, Cogni. Neurosci.* 33, 1275–1295. doi: 10.1080/23273798.2018.1476727

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

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., and Tanenhaus, M. K. (2014). Is it or isn't it: listeners make rapid use of prosody to infer speaker meanings. *Cognition* 133, 335–342. doi: 10.1016/j.cognition.2014.05.017

Kutlu, E. (2020). Now you see me, now you mishear me: Raciolinguistic accounts of speech perception in different English varieties. *J. Multiling. Multicult. Dev.* 42, 1–15. doi: 10.1080/01434632.2020.1835929

Langlois, V. J., and Arnold, J. E. (2020). Print exposure explains individual differences in using syntactic but not semantic cues for pronoun comprehension. *Cognition* 197:104155. doi: 10.1016/j.cognition.2019.104155

Laturnus, R. (2018). Perceptual Adaptation to Non-Native Speech: The Effects of Bias, Exposure, and Input Variation (Dissertation, New York University). Available at: http://search.proquest.com/docview/2081874820/abstract/6967135255D945DDPQ/1 (Accessed February 1, 2021).

Lee, J. W., and Jenks, C. J. (2019). Aestheticizing language: Metapragmatic distance and unequal Englishes in Hong Kong. *Asian Eng.* 21, 128–141. doi: 10.1080/13488678.2018.1448962

Lev-Ari, S. (2015). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Front. Psychol.* 5:1546. doi: 10.3389/fpsyg.2014.01546

Lev-Ari, S., and Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *J. Exp. Soc. Psychol.* 46, 1093–1096. doi: 10.1016/j.jesp.2010.05.025

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

MacCallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychol. Methods* 7, 19–40. doi: 10.1037/1082-989X.7.1.19

Marian, V., and Hayakawa, S. (2021). Measuring bilingualism: The quest for a bilingualism quotient. *Appl. Psycholinguist.* 42, 527–548. doi: 10.1017/S0142716420000533

Martin, T. (2017). *The Affirmative Action Empire*. Ithaca, New York: Cornell University Press.

Mathur, G., and Rathmann, C. (2009). "Variability in verbal agreement forms Across four signed languages," in *Laboratory Phonology 8*. eds. L. Goldstein, D. H. Whalen and C. T. Best (Berlin: De Gruyter, Inc.).

Namboodiripad, S. (2021). "Non-Dravidian elements and (non)diasystematic change in Malayalam," in *Constructions in Contact 2: Language Change, Multilingual Practices, and Additional Language Acquisition*. eds. S. Höder, and H. Boas (Benjamins).

Neidle, C. J., Kegl, J., Bahan, B., MacLaughlin, D., and Lee, R. G. (2000). *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, Massachusetts: MIT Press.

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., and Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Front. Psychol.* 6:312. doi: 10.3389/fpsyg.2015.00312

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., and Bunting, M. F. (2014). Clearing the garden-path: improving sentence processing through cognitive control training. *Language, Cogn. Neurosci.* 29, 186–217. doi: 10.1080/01690965.2012.758297

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.1253751

Ortega, L. (2020). The study of heritage language development From a bilingualism and social justice perspective. *Lang. Learn.* 70, 15–53. doi: 10.1111/lang.12347

Otheguy, R., García, O., and Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Appl. Lingu. Review* 6, 281–307. doi: 10.1515/applirev-2015-0014

Ou, J., Yu, A. C. L., and Xiang, M. (2021). Individual differences in categorization Gradience As predicted by online processing of phonetic cues During spoken word recognition: evidence From eye movements. *Cogn. Sci.* 45:e12948. doi: 10.1111/cogs.12948

Patrick, G. Z. (1989). *Roots of the Russian Language: An Elementary Guide to Word Building*. New York, NY: Passport Books.

Pinget, A.-F., Kager, R., and Van de Velde, H. (2020). Linking variation in perception and production in sound change: evidence from Dutch obstruent devoicing. *Lang. Speech* 63, 660–685. doi: 10.1177/0023830919880206

Quer, J., and Steinbach, M. (2019). Handling sign language data: The impact of modality. *Front. Psychol.* 10:483. doi: 10.3389/fpsyg.2019.00483

Ramjattan, V. A. (2019). The white native speaker and inequality regimes in the private English language school. *Intercult. Educ.* 30, 126–140. doi: 10.1080/14675986.2018.1538043

Rampton, M. B. H. (1990). Displacing the 'native speaker': expertise, affiliation, and inheritance. *ELT J.* 44, 97–101. doi: 10.1093/eltj/44.2.97

Rochat, P. (2010). What is really wrong with a priori claims of universality? Sampling, validity, process level, and the irresistible drive to reduce. *Behav. Brain Sci.* 33:107. doi: 10.1017/S0140525X10000233

Rosa, J. D. (2016). Standardization, racialization, Languagelessness: Raciolinguistic ideologies across communicative contexts. *J. Lingu. Anthropology* 26, 162–183. doi: 10.1111/jola.12116

Rosa, J., and Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Lang. Soc.* 46, 621–647. doi: 10.1017/S0047404517000562

Rothman, J., and Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Appl. Linguis.* 35, 93–98. doi: 10.1093/applin/amt049

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Res. High. Educ.* 33, 511–531. doi: 10.1007/BF00973770

Schouten, A. (2009). The critical period hypothesis: support, challenge, and reconceptualization. *Stud. Appl. Ling. TESOL* 9, 1–16. doi: 10.7916/salt.v9i1.1462

Sedarous, Y., and Namboodiripad, S. (2020). Using audio stimuli in acceptability judgment experiments. *Lang Ling Compass* 14:e12377. doi: 10.1111/lnc3.12377

Slezkine, Y. (1994). The USSR as a communal apartment, or how a socialist state promoted ethnic particularism. *Slav. Rev.* 53, 414–452. doi: 10.2307/2501300

Solís-Barroso, C., and Stefanich, S. (2019). Measuring language dominance in early Spanish/English bilinguals. *Language* 4:62. doi: 10.3390/languages4030062

Sorace, A. (2004). Native language attrition and developmental instability at the syntax-discourse interface: data, interpretations and methods. *Bilingualism: Lang. Cogn.* 7, 143–145. doi: 10.1017/S1366728904001543

Stern, H. H. (1983). *Fundamental Concepts of Language Teaching*. Cambridge, UK: Oxford University Press.

Sung, K.-Y. (2011). Factors influencing Chinese language learners' strategy use. *Int. J. Multiling.* 8, 117–134. doi: 10.1080/14790718.2010.532555

Surrain, S., and Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Lang. Cogn.* 22, 401–415. doi: 10.1017/S1366728917000682

Tanner, D., Mclaughlin, J., Herschensohn, J., and Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Lang. Cogn.* 16, 367–382. doi: 10.1017/S1366728912000302

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Washington, DC: American Psychological Association.

Thráinsson, H. (2012). "Ideal speakers and other speakers: The case of dative and Some other cases," in *Variation in Datives*. eds. B. Fernandez and R. Etxepare (Cambridge, UK: Oxford University Press), 161–188.

Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *J. Phon.* 71, 249–267. doi: 10.1016/j.wocn.2018.09.004

Tomoschuk, B., Ferreira, V. S., and Gollan, T. H. (2019). When a seven is not a seven: self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Lang. Cogn.* 22, 516–536. doi: 10.1017/S1366728918000421

van der Wouden, T., Hoekstra, H., Moortgat, M., Renmans, B., and Schuurman, I. (2002). "Syntactic Analysis in the Spoken Dutch Corpus (CGN)." in *Proceedings of the Third International Conference on Language Resources and Evaluation*; eds. M. González Rodríguez and C. Paz Suárez Araujo (European Language Resources Association (ELRA)), 768-773.

Vaughn, C., Baese-Berk, M., and Idemaru, K. (2019). Re-examining phonetic variability in native and non-native speech. *Phonetica* 76, 327–358. doi: 10.1159/000487269

Vinson, J. (1889). "Créoles," in *Dictionnaire Des Sciences Anthropologiques*. ed. A. Bertillon (Paris: Doin), 345–347.

Voeten, C. C. (2020). Individual differences in the adoption of sound change. *Lang. Speech* 25, 1–37. doi: 10.1177/0023830920959753

Williams, K. T. (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. New York, NY: Routledge.

Young, M. E. (2016). The problem with categorical thinking by psychologists. *Behav. Process.* 123, 43–53. doi: 10.1016/j.beproc.2015.09.009

Yu, A. C. L., and Zellou, G. (2019). Individual differences in language processing: phonology. *Annu. Review Ling.* 5, 131–150. doi: 10.1146/annurev-linguistics-011516-033815

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 October 2021
doi: 10.3389/fpsyg.2021.717352

Deconstructing the Native Speaker: Further Evidence From Heritage Speakers for Why This Horse Should Be Dead!

Wintai Tsehaye[1]†, Tatiana Pashkova[2]*†, Rosemarie Tracy[1] and Shanley E. M. Allen[2]*

[1] Department of English Linguistics, University of Mannheim, Mannheim, Germany, [2] Center for Cognitive Science, University of Kaiserslautern, Kaiserslautern, Germany

The category "native speaker" is flawed because it fails to consider the diversity between the speaker groups falling under its scope, as highlighted in previous literature. This paper provides further evidence by focusing on the similarities and differences between heritage speakers (HSs) and monolingually-raised speakers (MSs) of their heritage and majority languages. HSs are bilinguals who acquire a family (heritage) language and a societal (majority) language in early childhood. Naturalistic exposure from early childhood qualifies them as native speakers of their heritage language. Some HSs are simultaneous bilinguals, which makes them native speakers of their majority language as well. Others are early second language acquirers who may be indistinguishable from simultaneous bilinguals. Previous research shows that the heritage language productions of German HSs in the United States do not completely overlap with those of German MSs, who are, by default, native speakers. In overall clause type selection (independent main, coordinate main, and subordinate), the HSs differ from German MSs in German but are similar to English MSs in English. The present study examines the distribution of finite subordinate clauses and their types (relative, complement, and adverbial) across registers in 27 adolescent HSs of German in the United States, compared to 32 adolescent MSs of German and 32 MSs of English. All participants described a short video in two settings (formal/informal) and two modes (spoken/written). Results demonstrate that, even with respect to a specific phenomenon (subordinate clauses), HSs show similarities and differences to MSs of both languages. Concerning the distribution of subordinate clause types, HSs behave similarly to both English and German MSs. Concerning subordinate clauses in general, HSs use them less frequently than MSs in German. In English, the difference is more nuanced: HSs differentiate between settings in both modes, while MSs do so only in the written mode. This indicates that the category "native speaker" is not a meaningful descriptor since it covers speakers with varying production patterns. We propose that studies including native speakers should assure transparency and replicability of research by specifying and taking into account speaker characteristics such as bilingualism, proficiency, exposure and dominance.

Keywords: native speakers, heritage speakers, subordinate clauses, heritage German, majority language

OPEN ACCESS

Edited by:
Valentin Vulchanov, Norwegian University of Science and Technology, Norway

Reviewed by:
Maria M. Arredondo, University of Texas at Austin, United States
Jason Rothman, Arctic University of Norway, Norway

****Correspondence:***
Wintai Tsehaye
wtsehaye@mail.uni-mannheim.de
Tatiana Pashkova
pashkova@rhrk.uni-kl.de

† These authors have contributed equally to this work and share first authorship

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

Received: *30 May 2021*
Accepted: *13 September 2021*
Published: *05 October 2021*

Citation:
Tsehaye W, Pashkova T, Tracy R and Allen SEM (2021) Deconstructing the Native Speaker: Further Evidence From Heritage Speakers for Why This Horse Should Be Dead! Front. Psychol. 12:717352. doi: 10.3389/fpsyg.2021.717352

INTRODUCTION

The category "native speaker" has been used to characterize a particular speaker population for many years (see Hopp, 2016; Azar et al., 2019; Ionin et al., 2021; Redl et al., 2021 as recent cases in point). What most researchers seem to agree on is that a native speaker is defined as a speaker who acquires their language naturalistically in early childhood (Cook, 1999; Davies, 2004, 2013). Despite its popularity, this definition can be questioned. It has been criticized for being a political and ideological construct (Bonfiglio, 2010; Dewaele, 2018) and for discrediting late second language (L2) speakers as "deficient versions of natives" (Cook, 2016, p. 186). Another point of criticism is that the category is underspecified because it does not reflect the variation within the subgroups under its scope (Davies, 2004; Lowe, 2020). This criticism holds for the specific native speaker population considered in the present study, namely heritage speakers (HSs). They are broadly defined as "bilinguals who have acquired a family (heritage language) and a majority societal language naturalistically in early childhood" (Pascual et al., 2012, p. 450). Therefore, they are native speakers of both of their languages (Montrul, 2016; Kupisch and Rothman, 2018) irrespective of them being simultaneous bilinguals or early L2 acquirers of the majority language (Rothman and Treffers-Daller, 2014, p. 96).

Comparisons of HSs with monolingually-raised speakers (MSs) reveal areas of difference and similarity (Montrul, 2016, p. 208). The similarities with MSs can be found in both their heritage language (Nagy, 2015; Nagy and Lo, 2019; Łyskawa and Nagy, 2020) and their majority language (Kupisch et al., 2014; Pashkova et al., in press). The differences also become apparent in both their languages (Rothman, 2007; Polinsky, 2018; Scontras et al., 2018 for the heritage language; Scontras et al., 2017; Polinsky, 2018; Paradis, 2019 for the majority language). It is important to mention that the differences are not clear-cut but rather gradient. For example, in a study on clause-type use across registers, we found that German HSs with majority English showed similar distributional patterns in their heritage German productions in independent main clauses and different patterns in coordinate main clauses and subordinate clauses, compared to German MSs (Pashkova et al., in press). These results illustrate a more nuanced difference in clause type productions of MSs and HSs in their heritage language. Taken together, these findings indicate that the category "native speaker" fails to adequately reflect the variation between the speaker groups who fall under its scope, in this case, HSs and MSs.

Consequently, if a linguistic study states that it examined a group of native speakers, we cannot be absolutely certain who these speakers were and if their individual patterns of language use were comparable. The native speaker group could comprise for example MSs, HSs, or late L2 acquirers who emigrated and whose first language (L1) is undergoing attrition. Unquestionably, these speakers use their native language differently. Thus, further specification of the category "native speaker" is necessary to ensure transparency and replicability of research.

In the current study, we continue to address similarities and differences between two groups of native speakers, namely HSs and MSs. Focusing on finite subordinate clauses (SCs), we investigate their general use and the use of their types (complement, adverbial, and relative) across registers. This structural spectrum offers a promising area of variation in the two native speaker sub-groups because it is located at the interface of syntax and discourse (Sorace, 2011).

On the syntactic level, mastery of SCs is a potential source of variation in heritage language due to the complexity of SCs and different word order constraints in SCs in HSs' heritage and majority language (Pashkova et al., in press). Regarding SC types, differences in acquisition timing, paths, and the language input may play a key role in their later production (Andreou et al., 2020a). Researchers have suggested different acquisition trajectories of subordinate clause types (Vasilyeva et al., 2008; Paradis et al., 2017). In heritage language contexts, HSs and MSs presumably have similar acquisition conditions during infancy and early childhood, which then start to diverge once exposure to the majority language increases (around preschool/kindergarten), and especially once formal schooling sets in. Hence, for the heritage language, we can expect that the earliest acquired SC types will be similar in HSs' and MSs' productions, while the later acquired types might show more variation. In the majority language, HSs might experience a delay in late-acquired phenomena but eventually catch up with MSs (Schulz and Grimm, 2019), so we expect, apart from timing, no pronounced qualitative differences between HSs and MSs.

On the discourse level, register awareness creates another source of variation in heritage language use since HSs might not have sufficient exposure to a similarly wide range of registers as MSs of the same language (Polinsky, 2018, pp. 323–324; Aalberse et al., 2019, p. 148). HSs usually experience their heritage language in informal settings, most likely in oral interactions with family members, and might not be as familiar with formal registers. On the other hand, they use their majority language in a greater variety of communicative situations, so they develop a nuanced register awareness comparable to that of MSs of the majority language. Our research has shown that HSs can transfer their register awareness from their majority language to the heritage language, at least while choosing between independent main, coordinate main, and subordinate clauses (Pashkova et al., in press) when all options are available, in principle. What is yet unclear is whether and how this register awareness will manifest itself in a larger speaker sample and within specific sub-domains, such as the use of SC types.

In comparing HSs and MSs in their use of SCs and their types, we will argue that applying the category "native speaker" as a cover term for both these groups obscures a meaningful description of the variation in their patterns of language use. We address this terminological difficulty and propose adding further specification to the category "native speaker," such as presence of bilingualism, to enhance transparency and replicability. We furthermore briefly explore other variables, such as proficiency, exposure and dominance as potential characteristics for specification.

THEORETICAL AND CONCEPTUAL BACKGROUND

The Native Speaker Spectrum

A native speaker has been defined as "a person who learns a language as a child and continues to use it fluently as a dominant language" (Richards and Schmidt, 2013, p. 386). Other characteristics include grammatical and appropriate usage of the native language, self-identification with the community where it is spoken, and intuitions about (un)grammatical structures in that language. Davies (2013) adds creative performance and the ability to translate and interpret into the native language to the list of native speaker characteristics.

However, within these (extra-)linguistic features included in native speaker definitions, only one is uncontroversial and straightforward, namely the childhood acquisition of their L1 (Cook, 1999, p. 187; Davies, 2003, p. 436). Many of the other features mentioned can also be found in L2 speakers: they can use their L2 fluently, grammatically, appropriately, and intuitively, and be creative performers and translators/interpreters. This is the first point of criticism of the category "native speaker": how helpful is the category to group people with similar patterns of language use if the majority of its defining features appears in non-native speakers' productions as well (Lowe, 2020, pp. 21–22)?

Beyond linguistic considerations of fluency, accuracy, and intuition, the category "native speaker" has also been criticized for being politically and ideologically charged. It is noted that being a native speaker is associated with power, language ownership, and even positive personality traits (Bonfiglio, 2010). Race, background, and identity play a role in deciding whether a speaker could be a member of the native speaker group. Holliday (2009) writes that a prototypical English native speaker is a white Anglo-Saxon from an English-speaking western country, and those who do not fit this image might be excluded from native speakerhood. Bonfiglio (2010, p. 12) argues that, in some cases, nativeness is judged based on the speaker's ethnic/immigrant family background and not their language, for instance, Turkish HSs in Germany might not be readily viewed as German native speakers, even though they grew up in Germany and acquired German as one of their L1s.

Monolinguals and Heritage Speakers on the Native Speaker Spectrum

Monolingual speakers are the least disputed speaker population subsumed under the category "native speaker" as they only acquire their L1 naturalistically. HSs, however, have not always been included in the group of native speakers (Polinsky and Scontras, 2020). On the one hand, this might be surprising because HSs fit the criterion of naturalistic acquisition from early childhood. Some researchers might have excluded HSs from native speakers since they equate nativeness with high proficiency and dominance instead of seeing it as a product of naturalistic L1 acquisition (Kupisch and Rothman, 2018). On the other hand, such a confusion is understandable since we do frequently see differences in HSs' heritage language productions compared to MSs'. This is, however, an insufficient criterion for excluding HSs from the native speaker continuum as they are not the only group that might differ from a prototypical, highly proficient monolingual native speaker. We also find these differences in MSs with limited experience with the standard language and in late L2 bilinguals who have migrated and shifted dominance to the L2 and are experiencing L1 attrition (Dewaele, 2018; Kupisch and Rothman, 2018).

If the differences between HSs' and MSs' productions are not due to HSs being non-native speakers, what could they be attributed to? Many researchers agree that differences in amount and quality of input play a very important role in the eventual outcomes of heritage language acquisition (Montrul, 2016, pp. 117–119; Kupisch and Rothman, 2018; Aalberse et al., 2019, pp. 146–149). These differences in input could lead to variation in heritage language productions, for example, case marking in heritage German (Yager et al., 2015; Zimmer, 2020), inflected infinitives in heritage Brazilian Portuguese (Rothman, 2007), or the encoding of motion events in heritage Turkish (Goschler et al., 2020). However, some areas of the heritage language still display substantial similarity with MSs' productions, for example, voice onset times in heritage Italian (Nagy, 2015), case morphology in heritage Polish, Russian, and Ukrainian (Łyskawa and Nagy, 2020), or use of classifiers in heritage Cantonese (Nagy and Lo, 2019).

Yet, it would be too simplistic to say that one domain of heritage language grammar and use would show only similarities to MSs' productions, while another domain would be likely to show only differences. Some areas show both differences and similarities with MSs' productions. For instance, Brehmer and Usanova (2015) report that verb placement in heritage Russian in Germany is different in SCs compared to monolingual Russian, with an increase in use of the verb in clause-final position, which would be an expected transfer from German. However, main clauses in heritage Russian do not feature more use of the verb in second position (V2, required in German) than those in monolingual Russian. Thus, verb placement in heritage Russian exhibits difference and similarities with monolingual Russian. In a similar vein, our own previous research demonstrated that clause type use across different registers in heritage German also shows a combination of differences and similarities with monolingual German. While independent main clauses are used in the same manner by both speaker groups, coordinate main and subordinate clauses exhibit variation: HSs prefer coordinate main clauses, while MSs choose subordinate clauses more frequently (Pashkova et al., in press).

Concerning HSs' majority language, their linguistic behavior in everyday interactions is oftentimes comparable to that of MSs, especially once HSs reach early adulthood (Paradis, 2019). For example, HSs have been reported to not have a foreign accent in their majority language (Kupisch et al., 2014). Further, Pashkova et al. (in press) found no evidence that German HSs use different clause type patterns across registers in their majority English, compared to English MSs—overall, both groups used more independent main clauses in the written mode, more coordinate main clauses in the spoken mode, and more subordinate clauses in the formal setting. However, there is experimental evidence

that HSs might exhibit more fine-grained differences to English MSs in their majority English, for instance in the release of final stops (Polinsky, 2018, pp. 141–144), grammaticality judgments of subject–verb agreement (Paradis, 2019), and scope assignment (Scontras et al., 2017).

Summing up, HSs are typically native speakers of both of their languages since they typically acquire both languages naturalistically in early childhood. This does not mean, however, that HSs' linguistic performance is identical to that of prototypical, highly proficient MSs. These two groups of native speakers show differences and similarities in the patterns of their language use. Therefore, we propose further specification of the category "native speaker" in order to reflect this variability. Our study illustrates that an important variable to specify is the presence of bilingualism; additional specifications can include proficiency, exposure, and dominance.

Subordinate Clauses

The use of SCs and their types across registers is complex in that the speaker requires both syntactic knowledge and register awareness to decide on the appropriateness of SCs according to communicative situations (as explained in section "Register Characteristics of Subordinate Clauses," SCs are often more preferred in formal contexts). As specified in the Interface Hypothesis, structures involving both syntactic and pragmatic choices are particularly open to variation in terms of acquisition timing and/or cross-linguistic influence (Sorace, 2011; Tsimpli, 2014), thus leading to potentially different patterns across different types of natives speakers. We thereby add subordinate clause choice to the phenomena considered in interface research, given that register is a part of pragmatics, a language-external component (Tsimpli, 2014, p. 301). In the following section, we will examine the syntactic mastery of SCs and register awareness in both speaker groups.

Syntactic Characteristics of Subordinate Clauses

Subordinate clauses in general

Syntactically, SCs have the following features (Diessel, 2004, p. 48): they are integrated in the matrix clause, they are dependent structures that are formally incomplete without the matrix clause, and they are part of the same processing and planning unit as the associated matrix clause. This last feature is one of the reasons why SCs have been associated with higher syntactic complexity than juxtaposed matrix clauses (Polinsky, 2008; Neary-Sundquist, 2017; Peristeri et al., 2017; Sánchez Abchi and De Mier, 2017; Housen et al., 2019). Syntactic complexity has been defined, among other things, as the extent to which language users resort to syntactic embedding and SCs or as a structure which requires more steps in the syntactic derivation (Housen et al., 2012; Sanfelici and Schulz, 2021). However, the direct link between SCs and syntactic complexity has also been questioned: several researchers reported that textual complexity correlated not with the number of SCs but rather with mean length of nominal phrases and clauses (Lu, 2011; Wiese et al., 2020; Wang and Tao, 2020). Overall, the evidence for high complexity of SCs appears conflicting. Nevertheless, if SCs reflect textual complexity to some extent, we would expect fewer SCs in HSs' productions in their heritage language compared to MSs of that language.

In addition to the general complexity of SCs across languages, different word order constraints in HSs' heritage and majority language might play a role in SC production. This study examines HSs of German with English as their majority language. German and English differ in SC word order: In finite clauses introduced by complementizers and relative pronouns, German canonically exhibits subject-object-verb (SOV) structure,[1] while English has subject-verb-object (SVO) structure. This typological mismatch between the two languages of HSs might make the production of SCs in German harder for HSs than for MSs due to higher cognitive load because of the inhibition of one structure in the bilingual mind—in this case, SVO (Abutalebi and Green, 2016). This may lead to avoidance of SCs in the German productions of HSs (see Pashkova et al., in press, for a more detailed discussion).

Subordinate clause types

This section focuses on the syntactic characteristics of SC types and on how they might contribute to the variation between HSs and MSs. We follow previous researchers (e.g., Beaman, 1984; Diessel, 2004; Thompson et al., 2007; Paradis et al., 2017; Andreou and Tsimpli, 2020) in subdividing finite SCs into three categories: complement, adverbial, and relative clauses. In the following, we describe each clause type in detail and provide an overview of their L1 acquisition patterns.

Complement clauses are SCs that function as arguments of a predicate in the matrix clause (e.g., *She saw that a car was coming.*) (Biber et al., 1999, p. 658; Diessel, 2004, p. 1; Noonan, 2007; Lust et al., 2015, p. 301). Some researchers have suggested that complement clauses emerge early in L1 acquisition (Vasilyeva et al., 2008; Paradis et al., 2017), one of the reasons proposed for this being that they are narrowly syntactic structures that only require the knowledge of verb complement selection patterns and no pragmatic skills in discourse management (Mastropavlou and Tsimpli, 2011; Andreou, 2015; Andreou et al., 2020a). In child HSs, the accurate repetition of complement clauses in a sentence repetition task at the ages of 8–12 was reported to be associated with the amount of exposure to the language between ages 0 and 3 and at the age of 6 (Andreou et al., 2020a). This suggests that there are crucial periods for the development of complement clauses that correlate with their production later on. Hence, in the heritage language, we would expect similar production patterns in HSs and MSs because they received similar input at an age when language exposure could affect their emergence.

Adverbial clauses are SCs that modify main clauses similarly to adverbs and adverbial adjuncts modifying a proposition (e.g., *While she was walking, she saw an accident*) (Diessel, 2004, p. 1; Thompson et al., 2007). Contrary to narrowly syntactic structures, adverbial clauses, along with relative clauses, involve the syntax–discourse interface because they rely on discourse and pragmatics and call for discourse management skills (Peristeri et al., 2017, pp. 5, 11; Andreou et al., 2020a,b). For this

[1]Unintroduced subordinate clauses require verb-raising into second position, as in main clauses. Those cases are also accounted for in this study, see section "Data Coding."

reason, it has been argued that adverbial clauses are acquired later than complement clauses. Moreover, in child HSs, the accurate repetition of adverbial clauses at the ages of 8–12 was shown to be influenced by current language exposure (Andreou et al., 2020a). This suggests that adverbial clause use might be a locus for greater variation between heritage language and monolingual productions due to differences in the speakers' current language exposure.

Relative clauses are SCs that modify a noun phrase (NP) (e.g., *A woman who was pushing a baby stroller was walking down the street*) (Andrews, 2007). They are characterized by a syntactic gap that is associated with a relative pronoun at their left periphery and requires as its antecedent the relativized constituent of the matrix clause (Biber et al., 1999, p. 608; Diessel, 2004, p. 117). Similar to adverbial clauses, relative clauses are also located at the syntax-discourse interface and require discourse management skills, i.e., the ability to determine what is needed for referent specification in particular contexts. Therefore, one might expect relative clauses to be more influenced by later exposure, hence leading to greater variation between HSs' heritage language productions and those of MSs.

In the current study, we investigate whether the suggested differences of the acquisition onset of SC types impacts their use in HSs who are older than those examined in previous research (Andreou et al., 2020a).

Register Characteristics of Subordinate Clauses

Register is a variety definable in terms of situational parameters such as participants, channel, purpose, spoken or written mode, and formality of communication (Biber and Conrad, 2001, p. 175). In this study, we operationalize formality as spoken or written communication with public institutions, and informality as spoken or written communication with friends and family. HSs normally do not have as frequent exposure to a variety of registers in their heritage language compared to MSs of that language (Polinsky, 2018, pp. 323–324; Aalberse et al., 2019, p. 148 for recent mention of this tendency). Since the use of the heritage language is mostly limited to interactions with family members and perhaps members of a heritage language community, HSs are usually expected to be more familiar with informal registers and less familiar with formal registers. At the same time, HSs' majority language typically follows a different trajectory: they use it in a wider range of communicative situations and thus develop formal and informal register repertoires comparable to those of MSs. It is an interesting question, then, how HSs approach formal registers in their heritage language: would they use language patterns from the informal registers of their heritage language or would they try to rely on the formal register patterns from their majority language? Schleppegrell and Colombi (1997) argued for the latter option: they showed that Spanish HSs used very similar clause types in academic essays in heritage Spanish and majority English, despite being unfamiliar with formal academic registers in their heritage language.

Our recent study (Pashkova et al., in press) identified a similar tendency: German HSs showed similar clause type patterns in formal and informal registers in heritage German and majority English, which we called "an underlying register awareness"—HSs were able to transfer their register awareness from their majority language to their heritage language. Crucially, HSs used similar clause type patterns in heritage German compared not only to majority English but also to monolingual German. This possibility of transfer appears viable when the heritage and majority languages have similar register-related language use of the phenomenon under scrutiny, as was the case for clause type use in German and English (in both languages, MSs preferred independent main clauses in the written mode, coordinate main clauses in the spoken mode, and subordinate clauses in the formal setting). It is as yet unclear if register awareness can be attested in a larger data sample and transferred to another phenomenon, such as SC types. However, it is important to note that similar patterns of SC use in heritage and monolingual German did not mean the same frequency of SCs—HSs still used overall fewer SCs than MSs, most likely due to the syntactic characteristics of SCs outlined above.

Subordinate clauses and their types show variation across registers, which makes them an interesting phenomenon to examine with respect to register-related linguistic behavior of HSs. For instance, Koch and Oesterreicher (2012) outlined syntactic features of the language of immediacy, i.e., spontaneous face-to-face dialogues between familiar speakers, and the language of distance, i.e., carefully planned interactions between strangers in the public sphere. The language of immediacy is characterized by parataxis, whereas the language of distance is associated with hypotaxis. Our previous study (Pashkova et al., in press) confirmed this claim: in both English and German, we found more SCs in formal registers, which were similar to the language of distance, than in informal registers, similar to the language of immediacy.

Subordinate clause types are also subject to register variation. In English, for example, Biber and Gray (2016, pp. 87–100) reported more complement and adverbial finite clauses in conversation than in academic writing, and more wh-relative clauses in academic writing than in conversation. Beaman (1984) showed that nominal and relative subordinations occur more often in spoken narratives than in written ones, while adverbial subordinations are more frequent in written productions. Even though these findings do not map directly on the registers examined in the current study (a formal report to the police vs. an informal message to a close friend), we can still expect a certain variation in SC type productions according to formality. Our data will serve as an addition to the research on register repertoires of HSs because, to the best of our knowledge, there has not been a study that focuses on the systematic analysis of SC types according to formality.

The Present Study

To address the gaps in the literature just discussed, we pursue the following research questions (RQs) concerning the use of SCs in HSs' productions. Based on findings from the literature, we also lay out hypotheses and predictions for each question.

RQ 1: Do HSs show similarities or differences in the use of SCs according to register in their majority language (English)

compared to English MSs and in their heritage language (German) compared to German MSs?

Hypothesis 1: Based on our previous study of clause type use in a smaller participant sample (Pashkova et al., in press), we expect HSs to show similarities to English MSs and to differ from German MSs due to syntactic complexity and SOV word order of German SCs.

Prediction 1: Comparing HSs' majority English to monolingual English, we expect to find similar frequencies of SCs in all registers. Comparing HSs' heritage German to monolingual German, we expect to find similar patterns across registers but overall fewer SCs in heritage German.

RQ 2: Do HSs show similarities or differences in the use of SC types (relative, complement, and adverbial) according to formality[2] in their majority language (English) compared to English MSs and in their heritage language (German) compared to German MSs?

Hypothesis 2: We expect HSs to show similarities with English MSs, and a combination of differences and similarities with German MSs due to the different acquisition periods of SC types.

Prediction 2.1: Comparing HSs' majority English to monolingual English, we expect to find similar frequencies of SC types across settings (formal/informal). Comparing HSs' heritage German to monolingual German, we expect to find similar frequencies of complement clauses but different frequencies of adverbial and relative clauses, since the latter two SC types are assumed to be acquired later than complement clauses.

Prediction 2.2: Concerning the heritage language, we also expect to observe larger differences between HSs and MSs in the formal setting since HSs are less familiar with formal registers and we have no previous evidence that they can transfer their register awareness from majority English to heritage German in the use of SC types.

MATERIALS AND METHODS

Participants

For this study we looked at 91 adolescents aged 14–18 years (mean age = 16.1, SD = 1.39, 50 females), with 32 in each of the monolingual groups and 27 in the heritage German group with English as their majority language.

1. HSs of German with majority language English (mean age = 15.6, SD = 1.58, 12 females)
2. MSs of German (mean age = 16.6, SD = 0.91, 19 females)
3. MSs of English (mean age = 16.1, SD = 1.49, 19 females).

The HSs of German grew up speaking German with at least one L1 German-speaking parent in the household (21 HSs had one German-speaking parent, five had two, and one participant provided no answer). All speakers were either born in the United States, or moved there before age two. They did not receive bilingual education, but may have participated in German "Saturday schools" or other German-speaking activities in the community. Speakers of established German "language islands" were excluded from the study. We defined monolinguals as speakers whose L1 was the only language spoken at home, but who might have acquired further languages through foreign language instruction.

German HSs were recruited in Boston, Massachusetts; Madison, Wisconsin; and St. Paul, Minnesota by contacting German organizations and institutions as well as via social media platforms. German MSs were recruited via contacting German high schools in Berlin. English MSs were recruited in the same cities as German HSs (and in Long Island, New York) via social media platforms or through personal contacts. The socio-economic status of HSs' families was slightly higher than that of English and German MSs (see **Supplementary Appendix A** for detailed information on parental education) due to the nature of our HS participant pool, which mostly consisted of professionals whose move to the United States was work-related.

The German and English productions of the HSs as well as those of the English MSs were elicited in the United States and those of German MSs in Germany. The data for this study is openly accessible via the Research Unit Emerging Grammars (RUEG) 0.4.0 corpus (Wiese et al., 2020). Both English and German productions of HSs were compared to the productions of MSs of the respective language.

Materials and Procedure

The data was collected using the Language Situations methodology (RUEG group, 2018; Wiese, 2020), which elicits controlled, comparable, and quasi-naturalistic productions across registers. Participants watched a short non-verbal video depicting a minor car accident and recounted what they saw, imagining themselves witnesses to the accident. The procedure was divided into two settings. In the formal setting, the elicitor was formally dressed and met with the participant in a room set up like an office. In the informal setting, the elicitor was casually dressed and met with the participant in a more relaxed setting, with snacks and beverages offered. In order to enhance an easy-going, comfortable atmosphere, the elicitor and the participant engaged in 10–15 min of task-unrelated conversation in the target language at the beginning of the informal session. The participant watched the video three times in total (twice in the first setting, once in the second setting) and was then asked to recount it in two different modes: spoken and written.

The formal recounting was operationalized as a voice message to a police hotline (spoken) and a witness report to the police (written), while the informal recounting comprised a WhatsApp voice message (spoken) and a WhatsApp text message (written) to a friend. The order of settings (formal/informal) and modes (spoken/written) was balanced across participants. The MSs completed all tasks in one session. The HSs completed the tasks in two sessions—one for their majority language (English) and one for their heritage language (German)—with an interval of 3–5 days in between to minimize priming effects. The order of language sessions was counterbalanced across participants. Upon completion of all the narrative tasks, the participants filled out

[2]Due to the small sample size of SC types, we decided to collapse the four registers (formal spoken, formal written, informal spoken, informal written) into two formality conditions—formal vs. informal.

an online questionnaire[3] about their language background as well as a self-assessment of their abilities in each language. Self-assessment showed that HSs rated their speaking and writing skills higher in their majority English (speaking mean = 5, SD = 0; writing mean = 4.96, SD = 0.19) than in heritage German (speaking mean = 3.66, SD = 0.78; writing mean = 2.81, SD = 1.27). English monolinguals rated their skills comparably high (speaking mean = 4.75, SD = 0.51; writing mean = 4.53, SD = 0.57) to German monolinguals (speaking mean = 4.96, SD = 0.17; writing mean = 4.66, SD = 0.66).

Data Coding

As mentioned above, we investigated the use of SCs and their types (complement, adverbial, and relative) in narratives in English and German. In both languages, we examined only clauses that contained finite verbs to constrain the nature of the question. Morphologically non-canonical clauses, i.e., deviations with respect to person and number agreement, were still included, since they do not affect the type that the clause is assigned to. Subordinations missing complementizers or relative pronouns were included because a large proportion of the data stems from spoken productions and omitting complementizer "that" or relative pronouns "who" and "which" (in English) is common in spoken productions (Biber and Conrad, 2001). Non-finite constructions, such as infinitives, present participles, and past participles were excluded. All narratives were split into finite clauses, and each clause was coded for being an SC or a matrix clause. In German, SCs mostly exhibited finite verb-final structures, with the exception of unintroduced complement clauses (see below).[4] Weil V2 clauses were not coded as SCs since *weil* has lost its status of a subordinator in those constructions (Antomo and Steinbach, 2010; Reis, 2013).

Each SC was coded for its type: complement, adverbial, or relative.[5] We included both verb and noun complement clauses in our analysis even though the majority of L1 acquisition literature focuses on verb complements. Noun complement clauses usually complement a certain set of nouns such as *question, thought, report*, *argument* (Biber et al., 1999, pp. 645–656), and therefore appeared quite rarely in our data due to the content of the video. Since there were not enough cases to group them into a separate category, they were collapsed with verb complement clauses. Verb complement clauses (1a) should not be confused with what follows multi-word discourse markers *I think, I mean, I don't know, you know*, which look like epistemic expressions. In order to differentiate a discourse marker from an epistemic expression, a complementizer test was applied: if a complementizer/wh-pronoun was present or could be added after the expression in question, the expression was not taken to be a discourse marker and, hence, the following part was annotated as a complement clause (1b). If a complementizer was absent and could not be added, the expression was taken to be a discourse marker with no complement clause (1c). Each clause in square brackets in (1) was counted as one complement clause.

(1) a. They weren't looking and then realized [a car was coming$_{complement}$] (USbi52FE_fwE)[6]

b. I don't know [what else happened$_{complement}$] (USbi50FD_isE)

c. And then these two cars came by and like I dunno$_{discoursemarker}$ they came to the intersection and the guy dropped his ball (USmo64FE_isE)

In complement clauses, German exhibits finite verb-final structures (2a), but also allows for canonical V2 structures, if the complementizer is omitted after verbs of saying and thinking (2b). Each clause in square brackets in (2) was counted as one complement clause.

(2) a. und konnte daher nicht wissen [ob nach der Ball ein Mensch kommen würde$_{complement}$]

(USbi64MD_fwD)

"And due to this (the driver) could not know if a person would come after the ball."

b. Ich hoffe [ich konnte ihnen behilflich sein$_{complement}$]![7] (DEmo54FD_fwD)

"I hope I could be of help to you!"

All types of adverbial clauses (e.g., temporal, locative, causative, conditional, concessive) were put into one category. Each clause in square brackets in (3) was counted as one adverbial clause.

(3) a. I witnessed the crash [as I was walking along the side of a street$_{adverbial}$] (USbi55FD_fwE)

b. The car stopped short [because there was a dog trying to get the ball$_{adverbial}$] (USmo59FE_iwE)

c. [Als sie die straße überqueren wollten$_{adverbial}$], ist der Mann den Ball aus dem Hand gefallen.

(USbi64MD_fwD).

"As they wanted to cross the street, the ball dropped out of the man's hand."

As for relative clauses, we included not only those modifying an NP (4a,b) but also those modifying an entire proposition (4c,d) (Biber et al., 1999, p. 867). The reasoning here was similar to the inclusion of noun complement clauses: even though the majority of L1 acquisition literature focuses on NP-modifying relative clauses, there were a few cases of proposition-modifying relative clauses, which were, however, not numerous enough to form their own category, so they were collapsed with NP-modifying relative clauses. Even though there has been extensive research on different types of relative clauses in HSs (e.g., Polinsky, 2011; Albirini and Benmamoun, 2014), we did not distinguish between object and subject relative clauses because

[3]Questionnaire for adolescent participants of the Research Unit Emerging Grammars; https://osf.io/qhupg/

[4]We also included seven non-canonical V2 clauses clearly conceptualized as SCs: three complement clauses, two adverbial clauses, two relative clauses. We did not conduct a separate analysis V2 SCs due to their low frequency.

[5]We did not conduct fine-grained qualitative analyses of SC types such as examining word order, choice of complementizers or verb placement, although these characteristics are definitely worth exploring in further research. We did so since any further subdivision on the data would result in a too low number of data points in each subcategory to conduct a statistical analysis.

[6]The participant code in the examples includes the following information: US/DE, country of elicitation, United States or Germany; bi/mo, bilingual/monolingual speaker; 01, speaker number; M/F, speaker's sex; D/E, HS's heritage language (Deutsch for German) or monolinguals' L1 (English or German); f/i, formal/informal setting; s/w, spoken/written mode; D/E, language of elicitation, D for German or E for English.

[7]We preserved the original orthography of the written productions.

TABLE 1 | English clause productions by speaker group and register/formality.

Register	Formal spoken		Formal written		Informal spoken		Informal written	
Speaker group	HS	MS	HS	MS	HS	MS	HS	MS
All clauses	494	511	424	459	393	430	257	290
Subordinate clauses	145	128	119	144	88	95	58	50

Formality	Formal		Informal	
Speaker group	HS	MS	HS	MS
Complement clauses	41	49	40	44
Adverbial clauses	105	114	55	49
Relative clauses	118	109	51	52

TABLE 2 | German clause productions by speaker group and register/formality.

Register	Formal spoken		Formal written		Informal spoken		Informal written	
Speaker group	HS	MS	HS	MS	HS	MS	HS	MS
All clauses	448	732	358	625	370	638	219	399
Subordinate clauses	77	201	90	178	51	114	15	69

Formality	Formal		Informal	
Speaker group	HS	MS	HS	MS
Complement clauses	74	138	19	53
Adverbial clauses	22	65	23	69
Relative clauses	71	176	24	61

we did not have sufficient data points to perform a separate comparison of the two types. Each clause in square brackets in (4) was counted as one relative clause.

(4) a. it tried to like stop for this dog [that was running into the street$_{relative}$] (USmo65FE_isE)

b. Ein Mann [der anscheinend mit seiner Frau spazieren war$_{relative}$] prellte einen Fußball.

(DEmo69MD_fwD)

"A man who was walking apparently with his wife bounced a soccer ball."

c. The dog saw the ball and ran for it, [which caused the car in the front to stop$_{relative}$].

(USbi51FD_fwE)

d. und is dem ersten auto dann raufgefahren [was zu dem unfall geführt hat$_{relative}$]

(DEmo65FD_fsD)

"and drove into the first car which lead to the accident"

Tables 1, **2** show the total number of clause productions in English and German respectively.

Data Analysis

First, the data was coded for SCs and matrix clauses, resulting in a dependent variable "Clause type" with two levels (1 for SC and 0 for matrix clause). We analyzed the use of SCs vs. matrix clauses using generalized binomial linear mixed effect models in R (R Core Team, 2021) and the lme4 package (Bates et al., 2015). We maximally specified the fixed effects by including all theoretically relevant independent variables and their interactions: bilingualism (heritage bilingual/monolingual), setting (formal/informal), mode (spoken/written). We contrast-coded the factors using sum contrast coding (–0.5/0.5). We attempted to maximally specify the random effect of participants and included the random slopes for setting and mode (Barr et al., 2013). The maximal specification worked for German SCs, but not for English SCs, where it led to overfitting, so we removed the random slopes and left only the random intercept.

Second, each SC was coded for its type, resulting in a dependent variable "SC type" with three levels (complement, adverbial, and relative). Then, we recoded the dependent variable "SC type" into three separate dependent variables "Complement clause", "Adverbial clause", and "Relative clause" with two levels (1 and 0). After this manipulation, each SC type was analyzed independently from the other two types also using generalized binomial linear mixed effect models. Due to the small sample size of each SC type (**Tables 1**, **2**), we collapsed the spoken and written modes within each setting and only included the independent variables of bilingualism (heritage bilingual/monolingual) and setting (formal/informal) and their interaction. We contrast-coded the factors using sum contrast coding (–0.5/0.5). Where possible, we maximally specified the random effect of participants by including the random slopes for setting. If this led to a perfect correlation of fixed effects or a random effect variance estimated at 0 or 1, we removed the random slope. In the next section, we report the z- and p-values of the models, for full model summaries, see **Supplementary Appendix B**.

RESULTS

Majority and Monolingual English

Subordinate Clauses in English

For English SCs, we observed a main effect of setting ($z = 4.70$, $p \leq 0.001$): speakers produced more SCs in the formal setting more than in the informal setting (**Figure 1**). In addition, we observed a three-way interaction between bilingualism, setting, and mode ($z = 2.02$, $p = 0.043$). To interpret this interaction, we ran separate models for HSs and MSs. HSs showed a main effect of setting ($z = 2.71$, $p = 0.007$), while MSs showed a main effect of setting ($z = 4.04$, $p \leq 0.001$) and an interaction between setting and mode ($z = -2.46$, $p = 0.014$). Tukey's multiple comparison test (MCT, run with *emmeans* package, Lenth, 2021) revealed a significant difference between the formal and the informal settings in the written mode (estimate = –0.51, SE = 0.16, $z = -3.26$, $p = 0.006$) and an absence of such a difference in the spoken mode (estimate = 0.19, SE = 0.16, $z = 1.24$, $p = 0.602$). This shows that German HSs and English MSs partially overlapped in their SC productions. While they behaved similarly in the written mode, they diverged in the spoken mode: HSs distinguished between the settings whereas MSs did not. Additionally, for both speaker groups, setting played a key role in SC production.

Subordinate Clause Types in English

For English complement clauses, we observed a main effect of setting ($z = -3.73$, $p \leq 0.001$): there were fewer complement

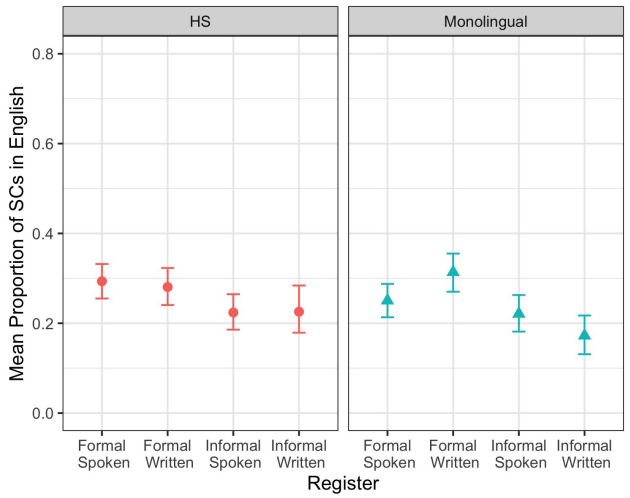

FIGURE 1 | Mean proportion of SCs in English by speaker group and register.

clauses in the formal setting than in the informal one (**Figure 2A**). For English adverbial clauses and relative clauses, we did not observe any main effects or interactions (**Figures 2B,C**). These results indicate that German HSs and English MSs performed similarly regarding the production of all SC types, and formality played a role only for complement clauses, with fewer complement clauses in the formal setting.

Heritage and Monolingual German

Subordinate Clauses in German

For German SCs, we observed two main effects and two interactions. First, there was a main effect of bilingualism ($z = -3.55$, $p \leq 0.001$), with HSs producing fewer SCs than MSs (**Figure 3**). Second, we found a main effect of setting ($z = 6.35$, $p \leq 0.001$): there were more SCs in the formal setting than in the informal setting. Then, we observed an interaction of setting and mode ($z = -2.98$, $p = 0.003$), with a greater difference between the formal and informal settings in the written mode (estimate = 1.08, SE = 0.18, $z = 5.94$, $p \leq 0.001$) than in the spoken mode (estimate = 0.45, SE = 0.13, $z = 3.37$, $p = 0.004$), according to Tukey's MCT. Finally, we observed a three-way interaction between bilingualism, setting, and mode. To interpret it, we ran separate models for HSs and MSs. The HS model indicated a main effect of setting ($z = 4.61$, $p \leq 0.001$), with more SCs in the formal setting than in the informal setting. In addition, there was an interaction of setting and mode. Tukey's MCT revealed a difference between the formal and informal setting in the written mode (estimate = 1.45, SE = 0.30, $z = 4.84$, $p \leq 0.001$) but not in the spoken mode (estimate = 0.22, SE = 0.20, $z = 1.09$, $p = 0.698$). The MS model showed only a main effect of setting ($z = 4.36$, $p \leq 0.001$). This shows that German HSs and MSs differed in the overall SC productions: while HSs distinguished between the settings only in the written mode, MSs did so in both modes. In addition, for both speaker groups, setting played a key role in SC production.

Subordinate Clause Types in German

For German complement clauses, we observed a main effect of setting ($z = -5.74$, $p \leq 0.001$), with fewer complement clauses in the formal setting than in the informal setting (**Figure 4A**). For adverbial clauses, we observed a main effect of setting ($z = 2.90$, $p = 0.004$), with more adverbial clauses in the formal setting than the informal setting (**Figure 4B**). For relative clauses, we observed a main effect of setting ($z = 2.30$, $p = 0.022$), with more relative clauses in the formal setting than the informal setting (**Figure 4C**). These results indicate that German HSs and German MSs performed similarly regarding the production of all SC types. Formality played a role for both speaker groups: they produced fewer complement clauses but more adverbial clauses and relative clauses in the formal setting than the informal setting.

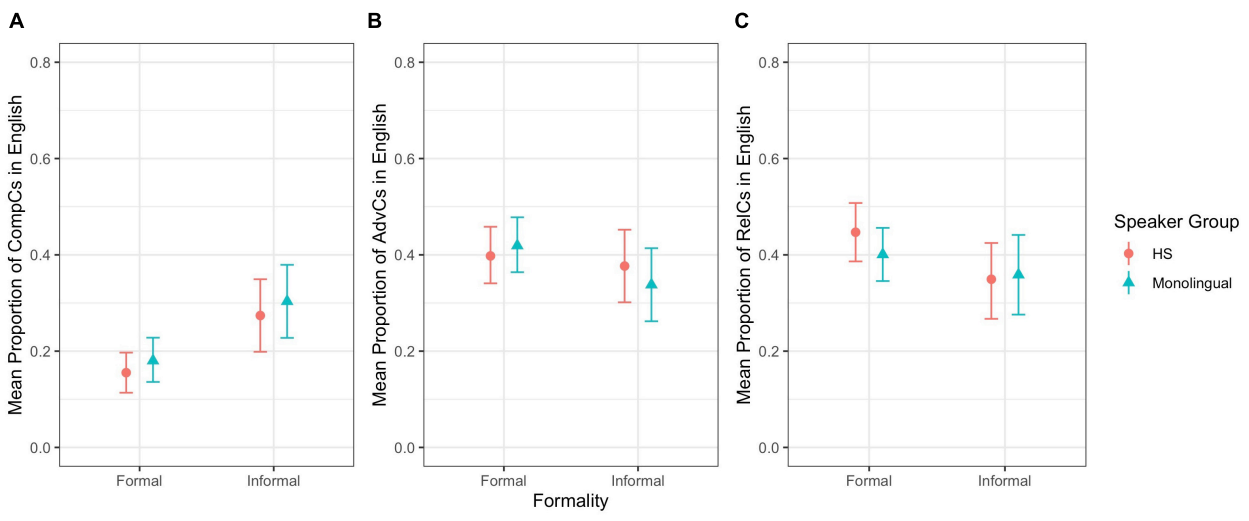

FIGURE 2 | Mean proportion of SC types in English by speaker group and formality: **(A)** Complement Clauses, **(B)** Adverbial Clauses, **(C)** Relative Clauses.

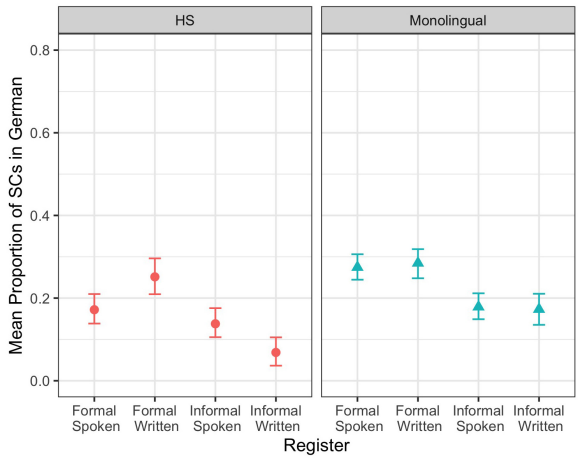

FIGURE 3 | Mean proportion of SCs in German by speaker group and register.

DISCUSSION

This study aimed at presenting reasons for why the category "native speaker" is flawed and should be further specified to account for the variation between the groups that fall under its scope. Such a specification would enhance transparency and replicability of research. We analyzed two native speaker groups—HSs and MSs—to argue that there are differences and similarities, as well as a combination of both, between the groups. In particular, we compared German HSs residing in the United States with English and German MSs. We looked at the use of SCs and their types (complement, adverbial, and relative) in spoken and written narratives across registers.

Our first research question focused on whether HSs use finite SCs in a similar or different way in their majority language compared to English MSs and in their heritage language compared to German MSs. With respect to HSs' majority language, our data does not confirm Hypothesis 1 and Prediction 1, which state that in their majority language, German HSs will perform similarly to English MSs. Overall, both speaker groups produce more SCs in the formal setting, confirming previous results, thus exhibiting similarity (see Pashkova et al., in press). This similarity is however only partial because a closer look at SC productions across registers reveals that HSs distinguished between the settings in spoken and written modes while MSs did so only in the written mode. With respect to HSs' heritage language, our data confirms Hypothesis 1 and Prediction 1, which state that in their heritage language, German HSs will produce significantly fewer SCs than German MSs. Additionally, HSs distinguished between the settings only in the written mode, while MSs did so in both modes. This can be attributed to the cognitive load of spoken online productions in combination with the general complexity of SCs and word order differences in SCs in English and German (Pashkova et al., in press).

Our second research question zoomed in on the use of finite SC types according to formality. We wanted to know whether HSs would show similarities or differences in their majority language compared to English MSs and in their heritage language compared to German MSs. With respect to HSs' majority language, our data confirms Hypothesis 2 and Prediction 2.1, which state that HSs and MSs should show similar frequencies of SC types across settings. With respect to HSs' heritage language, our data does not confirm Hypothesis 2 and Prediction 2.1, which expect a combination of differences and similarities between HSs and MSs, because both speaker groups in fact behaved similarly regarding the frequencies of SC types across settings. Consequently, we did not find any support for Prediction 2.2, which argued for a bigger difference between HSs and MSs in the formal setting.

Overall, the results show that the locus of variation between HSs and MSs is not where we predicted it to be. For English

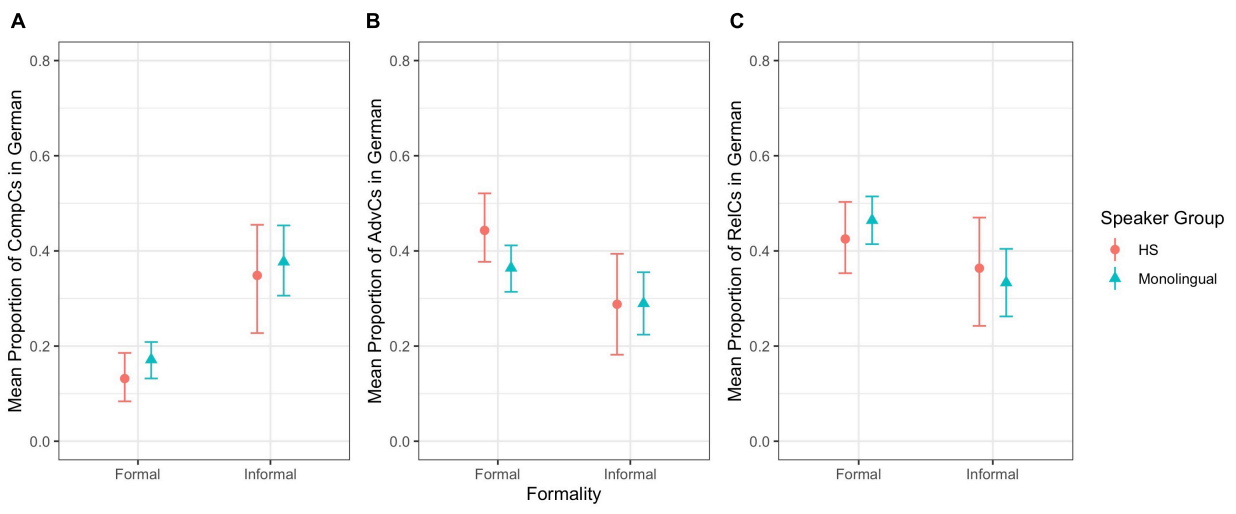

FIGURE 4 | Mean proportion of SC types in German by speaker group and formality: **(A)** Complement Clauses, **(B)** Adverbial Clauses, **(C)** Relative Clauses.

SCs, we expected to find only similarities between HSs and MSs, and instead we observed a combination of differences and similarities. HSs adhere to formality distinctions regardless of mode, unlike English MSs, who do so only in the written mode. This could be attributed to the different attitudes toward our study among HSs and MSs: HSs were well aware that their language competence was under scrutiny, and were probably trying to show their best language skills. This is especially true for the heritage language but could also have influenced their performance in the majority language, which might explain their strict adherence to the formality distinction in both modes. This illustrates that the two groups of native speakers show variation in their performance, potentially due to extralinguistic factors such as their perception of the situation. Therefore, the category "native speaker" groups together speakers with different patterns of language use and is not specific enough to allow comparability in a speaker population.

Another unpredicted result is that in German, HSs behave similarly to MSs with regard to all SC types, even adverbial and relative clauses, which we expected to differ between the speaker groups due to their later acquisition and location at the syntax-discourse interface. This is contrary to the previous findings by Andreou et al. (2020a), who showed that the current language exposure influences the production of adverbial clauses by child HSs in a sentence repetition task. However, their participants were much younger than ours (mean age 9.01 vs. mean age 15.6), which could be the reason for the discrepancy in our results. Perhaps, the use of adverbial clauses is influenced not only by the current language exposure but also by the overall cognitive maturity of the speaker (see Paradis et al., 2017 on the advantages of higher cognitive maturity in early L2 acquisition). Furthermore, the absence of difference could be attributed to the relatively small sample size in this study, which could have prevented us from capturing it. Productions of more speakers need to be analyzed to confirm our result. The analysis of SC types and SCs in German illustrated that we can still find similarities within a narrower phenomenon (SC types) between the sub-groups of native speakers even if a more general phenomenon (SCs) shows differences between the same speaker groups.

An additional unexpected finding was that concerning SC types, HSs behaved similarly to German MSs in their heritage language and similarly to English MSs in their majority language, even though the MSs of the respective languages behaved differently—in English, formality only had an effect on complement clauses, whereas in German, formality had an effect on all SC types. This shows that German and English differ in their formality-related language use and that HSs are able to adapt to the MS pattern in both their languages. This is surprising since the HSs' ability to adjust their SC type productions in their heritage language does not appear to originate from their exposure to formal registers in German or from transfer of their formality awareness from English into German. Further research is needed to pinpoint the source of this behavior.

The presented findings lead us to the conclusion that the category "native speaker" is too general to adequately define a speaker population because the speakers subsumed under this category may well differ in their linguistic behavior. Therefore, we argue for a more specific categorization, which provides more fine-grained information on their language background, allowing the possibility of capturing both group and individual variation, which are gradient (Ortega, 2020). Previous literature suggests that the category" "native speaker" should be replaced with "L1 user" (Dewaele, 2018). We argue for the necessity of further specification since even within L1 users, we can see differences as illustrated throughout this paper. This specification could include information on bilingualism, language exposure, proficiency, and dominance. In the current statistical analysis, we included only the variable of bilingualism in heritage language context. Further studies are needed to examine the influence of proficiency, language exposure, and dominance, which we expect to play a role in the variability among native speakers. Following this suggestion, for example, the majority of our German HSs could be described as bilinguals who are simultaneously raised in German and English, residing in the United States, with English as their current dominant language and German as their less dominant language. A typical German MS could be described as a monolingually-raised German speaker, residing in Germany, with German as their current dominant language.

One limitation of the present study, as already mentioned, is the relatively small sample size of the three SC types, which did not allow us to look into the interaction of bilingualism, formality and spoken/written mode. Since this interaction proved significant in the SC use, it would be very interesting to examine it in SC types as well. Due to a small sample size, we also were not able to assess potential qualitative differences in SC types (word order, choice of complementizer, or verb placement). Another possible extension of the current study is to examine further heritage-majority language pairs, probably typologically more distant, to see whether the patterns we describe here would manifest themselves in other native speaker groups. The RUEG corpus, which provided the data analyzed in this study, is a useful resource for such an extension since it contains comparable data for Greek, Turkish, and Russian HSs in Germany and the United States, plus data for their monolingual counterparts. Another aspect that could be addressed in future studies is the register-related language use in English, German, and possibly other languages. It is noteworthy that English and German MSs in our study did not behave similarly with respect to formality, and further research would be needed to uncover the possible sources of this difference. An additional step could be the inclusion of a wider range of registers with the same formality and mode distinctions, to see whether the formality sensitivity is tied to a particular situation (e.g., a police report) or if it is more general.

CONCLUSION

This study investigated the appropriateness of the category "native speaker" by comparing productions of two native speaker groups, namely heritage and monolingual speakers. We assessed the use of SCs and their types (complement, adverbial, and relative) in narratives produced by adolescent HSs of German

in the United States in comparison with adolescent German and English MSs. We provided evidence that there are similarities, differences, and a combination of both in the productions of HSs and MSs. Our results show similarities in the production of SC types between HSs' majority English and monolingual English, as well as between heritage and monolingual German. Differences were found in SC productions between heritage and monolingual German. A combination of differences and similarities was found in SC productions between majority and monolingual English. These findings support existing criticism of the category "native speaker" and further highlight its underspecification. As is, the category fails to adequately reflect the variation among speaker groups who fall under its scope. Therefore, we argue that we should enhance the category "native speaker" with more specific descriptions of speaker groups in order to provide unambiguous information about them.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: RUEG corpus https://doi.org/10.5281/zenodo.3236068.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Deutsche Gesellschaft für Sprachwissenschaft ethics committee and the Institutional Review Board (IRB) University of Maryland at College-Park. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

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

FUNDING

The research results presented in this publication were funded by the German Research Foundation (DFG) as part of the research unit *Emerging grammars in language contact situations: a comparative approach* (FOR 2537) in projects P5 (project no. 394995401, GZ TR 238/5-1 and P2 (project no. 394837597, GZ AL 1886/2-1).

ACKNOWLEDGMENTS

We thank all project members and research assistants. The publication of this article was funded by the Baden-Württemberg Ministry of Science, Research and the Arts and by the University of Mannheim.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717352/full#supplementary-material

REFERENCES

Aalberse, S., Backus, A., and Muysken, P. (2019). *Heritage languages: A language contact approach*, Vol. 58. Amsterdam: Benjamins, doi: 10.1075/sibil.58

Abutalebi, J., and Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Biling. Lang. Cogn.* 19, 689–698. doi: 10.1017/S1366728916000225

Albirini, A., and Benmamoun, E. (2014). Aspects of second-language transfer in the oral production of Egyptian and Palestinian heritage speakers. *Int. J. Bilingual.* 18, 244–273. doi: 10.1177/1367006912441729

Andreou, M. (2015). *The Effects of Bilingualism on Verbal and Non Verbal Cognition: The Micro- and Macro-Structure of Narratives in the Weak and the Dominant Language of the Bilingual Child.* Ph.D thesis. Thessaloniki: Aristotle University of Thessaloniki. doi: 10.13140/RG.2.2.11670.29767

Andreou, M., and Tsimpli, I. M. (2020). "Bilingualism, biliteracy and syntactic complexity: The role of crosslinguistic influence and cognitive skills," in *Language Acquisition, Processing and Bilingualism: Selected Papers from the Romance Turn VII*, eds A. Cardinaletti, C. Branchini, G. Giusti, and F. Volpato (Cambridge: Cambridge Scholars Publishing), 132–159.

Andreou, M., Torregrossa, J., and Bongartz, C. (2020a). "Sentence Repetition Task as a Measure of Language Dominance," in *BUCLD 45 Proceedings*, (Boston: Boston University).

Andreou, M., Tsimpli, I. M., Durrleman, S., and Peristeri, E. (2020b). Theory of Mind, Executive Functions, and Syntax in Bilingual Children with Autism Spectrum Disorder. *Languages* 5:67. doi: 10.3390/languages5040067

Andrews, A. D. (2007). "Relative clauses," in *Language Typology and Syntactic Description*, ed. T. Shopen (Cambridge: Cambridge University Press), 206–236. doi: 10.1017/CBO9780511619434.004

Antomo, M., and Steinbach, M. (2010). Desintegration und interpretation: Weil-V2-sätze an der schnittstelle zwischen syntax, semantik und pragmatik. *Zeitschrift Fur Sprachwissenschaft* 29, 1–37. doi: 10.1515/ZFSW.2010.001

Azar, Z., Backus, A., and Özyürek, A. (2019). General- and Language-Specific Factors Influence Reference Tracking in Speech and Gesture in Discourse. *Discourse Proces.* 56, 553–574. doi: 10.1080/0163853X.2018.1519368

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beaman, K. (1984). "Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse," in *Coherence in Spoken and Written Discourse*, eds D. Tannen and R. Freedle (Cambridge: Cambridge Scholars Publishing), 45–80.

Biber, D., and Conrad, S. (2001). "Register Variation: A Corpus Approach," in *The Handbook of Discourse Analysis*, eds D. Schiffrin, D. Tannen, and H. E. Hamilton (New Jersey, NJ: Blackwell Publishers), 175–196.

Biber, D., and Gray, B. (2016). *Grammatical Complexity in Academic English*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511920776

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Pearson Educaiton Limited.

Bonfiglio, T. P. (2010). *Mother tongues and nations: The invention of the native speaker*. Berlin: De Gruyter.

Brehmer, B., and Usanova, I. (2015). "Let's fix it? Cross-linguistic influence in word order patterns of Russian heritage speakers in Germany," in *Transfer Effects in Multilingual Language Development*, ed. H. Peukert (Amsterdam: John Benjamins Publishing Company), 161–188. doi: 10.1075/hsld.4.08bre

Bulte, B., and Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *J. Second Lang. Writ.* 26, 42–65. doi: 10.1016/j.jslw.2014.09.005

Cook, V. (1999). Going beyond the Native Speaker in Language Teaching. *TESOL Quart.* 33:185. doi: 10.2307/3587717

Cook, V. (2016). Where Is the Native Speaker Now? *TESOL Quart.* 50, 186–189. doi: 10.1002/tesq.286

Davies, A. (2003). *The native speaker: Myth and reality*. Bristol: Multilingual Matters, doi: 10.21832/9781853596247

Davies, A. (2004). "The native speaker in applied linguistics," in *The Handbook of Applied Linguistics*, eds A. Davies and C. Elder (Oxford: Blackwell), 431–450. doi: 10.1002/9780470757000.ch17

Davies, A. (2013). *Native Speakers and Native Users*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9781139022316

Dewaele, J. M. (2018). Why the Dichotomy "L1 Versus LX User" is Better than "Native Versus Non-native Speaker.". *Appl. Linguist.* 39, 236–240. doi: 10.1093/applin/amw055

Diessel, H. (2004). *The Acquisition of Complex Sentences*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511486531

Goschler, J., Schroeder, C., and Woerfel, T. (2020). "Convergence in the encoding of motion events in heritage Turkish in Germany," in *Studies in Turkish as a Heritagae Language*, ed. F. Bayram (Amsterdam: John Benjamins Publishing Company), 87–103. doi: 10.1075/sibil.60.05gos

Holliday, A. (2009). "English as a lingua franca, 'non-native speakers' and cosmopolitan realities," in *English as an International Language: Perspectives and Pedagogical Issues*, ed. F. Sharifian (Bristol: Channel View Publications), 19–33. doi: 10.21832/9781847691231-005

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Lang. Res.* 32, 277–307. doi: 10.1177/0267658315624960

Housen, A., Kuiken, F., and Vedder, I. (2012). "Complexity, accuracy and fluency," in *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, eds A. Housen, F. Kuiken, and I. Vedder (Amsterdam: John Bejamins), 1–20. doi: 10.1075/lllt.32.01hou

Housen, A., De Clercq, B., Kuiken, F., and Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second Lang. Res.* 35, 3–21. doi: 10.1177/0267658318809765

Ionin, T., Choi, S. H., and Liu, Q. (2021). Knowledge of indefinite articles in L2-English: Online vs. offline performance. *Second Lang. Res.* 37, 121–160. doi: 10.1177/0267658319857466

Koch, P., and Oesterreicher, W. (2012). "Language of immediacy – language of distance: orality and literacy from the perspective of language theory and linguistic history," in *Communicative Spaces: Variation, Contact, Change. Papers in Honour of Ursula Schaefer,* eds C. Lange, B. Weber, and G. Wolf (Frankfurt: Peter Lang), 441–473.

Kupisch, T., and Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Bilingual.* 22, 564–582. doi: 10.1177/1367006916654355

Kupisch, T., Barton, D., Hailer, K., Klaschik, E., Stangen, I., Lein, T., et al. (2014). Foreign Accent in Adult Simultaneous Bilinguals. *Heritage Lang. J.* 11, 123–150. doi: 10.46538/hlj.11.2.2

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.6.1.* Available online at: https://CRAN.R-project.org/package=emmeans

Lowe, R. J. (2020). *Uncovering Ideology in English Language Teaching*. Berlin: Springer.

Lust, B., Foley, C., and Dye, C. D. (2015). "The first language acquisition of complex sentences," in *The Cambridge Handbook of Child Language*, eds E. L. Bavin and L. R. Naigles (Cambridge: Cambridge University Press), 298–323.

Łyskawa, P., and Nagy, N. (2020). Case Marking Variation in Heritage Slavic Languages in Toronto: Not So Different. *Lang. Learn.* 70, 122–156. doi: 10.1111/lang.12348

Mastropavlou, M., and Tsimpli, I. M. (2011). Complementizers and subordination in typical language acquisition and SLI. *Lingua* 121, 442–462. doi: 10.1016/j.lingua.2010.10.009

Montrul, S. (2016). *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9781139030502

Nagy, N. (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164, 309–327. doi: 10.1016/j.lingua.2014.04.012

Nagy, N., and Lo, S. (2019). Classifier use in Heritage and Hong Kong Cantonese. *Asia Pacific Lang. Variat.* 5, 84–108. doi: 10.1075/aplv.17001.nag

Neary-Sundquist, C. A. (2017). Syntactic complexity at multiple proficiency levels of L2 German speech. *Int. J. Appl.* Linguist. 27, 242–262. doi: 10.1111/ijal.12128

Noonan, M. (2007). "Complementation," in *Language Typology and Syntactic Description*, ed. T. Shopen (Cambridge: Cambridge University Press), 52–150. doi: 10.1017/CBO9780511619434.002

Ortega, L. (2020). The Study of Heritage Language Development From a Bilingualism and Social Justice Perspective. *Lang. Learn.* 70, 15–53. doi: 10.1111/lang.12347

Paradis, J. (2019). "English Second Language Acquisition from Early Childhood to Adulthood: The Role of Age, First Language, Cognitive, and Input Factors," in *Proceedings of the 43rd Boston University Conference on Language Development*, eds M. M. Brown and B. Dailey (Somerville, MA: Cascadilla Press), 11–26.

Paradis, J., Rusk, B., Duncan, T. S., and Govindarajan, K. (2017). Children's Second Language Acquisition of English Complex Syntax: The Role of Age, Input, and Cognitive Factors. *Annu. Rev. Appl. Linguist.* 37, 148–167. doi: 10.1017/S0267190517000022

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

Peristeri, E., Andreou, M., and Tsimpli, I. M. (2017). Syntactic and Story Structure Complexity in the Narratives of High- and Low-Language Ability Children with Autism Spectrum Disorder. *Front. Psychol.* 8:02027. doi: 10.3389/fpsyg.2017.02027

Polinsky, M. (2008). Gender under incomplete acquisition: heritage speakers' knowledge of noun categorization. *Herit. Lang. J.* 6, 40–71. doi: 10.46538/hlj.6.1.3

Polinsky, M. (2011). Reanalysis in Adult Heritage Language: New evidence in support of attrition. *Stud. Second Lang. Acquisit.* 33, 305–328. doi: 10.1017/S027226311000077X

Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press, doi: 10.1017/9781107252349

Polinsky, M., and Scontras, G. (2020). A roadmap for heritage language research. *Bilingual. Lang. Cognit.* 23, 50–55. doi: 10.1017/S1366728919000555

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Redl, T., Frank, S. L., de Swart, P., and de Hoop, H. (2021). The male bias of a generically-intended masculine pronoun: Evidence from eye-tracking and sentence evaluation. *PLoS One* 16:e0249309. doi: 10.1371/journal.pone.0249309

Reis, M. (2013). Weil-V2"-Sätze und (k)ein Ende? Anmerkungen zur Analyse von Antomo & Steinbach (2010). *Zeitschrift Fur Sprachwissenschaft* 32, 221–262. doi: 10.1515/zfs-2013-0008

Richards, J. C., and Schmidt, R. W. (2013). *Longman Dictionary of Language Teaching and Applied Linguistics*, 4th Edn. London: Routledge, doi: 10.4324/9781315833835

Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *Int. J. Bilingual.* 11, 359–389. doi: 10.1177/13670069070110040201

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

RUEG group (2018). *RUEG Language Situations Elicitation 2018/19*. Charlottesville,VA: OSF.

Sánchez Abchi, V., and De Mier, V. (2017). Syntactic complexity in narratives written by Spanish heritage speakers. *Vigo Int. J. Appl. Linguist.* 14, 125–148.

Sanfelici, E., and Schulz, P. (2021). Can frequency account for the grammatical choices of children and adults in nominal modification contexts? Evidence from elicited production and child-directed speech. *Languages* 6, 1–31. doi: 10.3390/languages6010035

Schleppegrell, M. J., and Colombi, M. C. (1997). Text Organization by Bilingual Writers. *Written Commun.* 14, 481–503.

Schulz, P., and Grimm, A. (2019). The age factor revisited: Timing in acquisition interacts with age of onset in bilingual acquisition. *Front. Psychol.* 9:02732. doi: 10.3389/fpsyg.2018.02732

Scontras, G., Polinsky, M., and Fuchs, Z. (2018). In support of representational economy: Agreement in heritage Spanish. *Glossa J. General Linguist.* 3:1. doi: 10.5334/gjgl.164

Scontras, G., Polinsky, M., Tsai, C.-Y. E., and Mai, K. (2017). Cross-linguistic scope ambiguity: When two systems meet. *Glossa J. General Linguist.* 2, 1–28. doi: 10.5334/gjgl.198

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguist. Approaches Bilingual.* 1, 1–33. doi: 10.1075/lab.1.1.01sor

Thompson, S. A., Longacre, R. E., and Hwang, S. J. J. (2007). "Adverbial clauses," in *Language Typology and Syntactic Description, Second Edition: Volume II: Complex Constructions*, ed. T. Shopen (Cambridge: Cambridge University Press), 237–300. doi: 10.1017/CBO9780511619434.005

Tsimpli, I. M. (2014). Early, late or very late? *Linguist. Approac. Bilingual.* 4, 283–313. doi: 10.1075/lab.4.3.01tsi

Vasilyeva, M., Waterfall, H., and Huttenlocher, J. (2008). Emergence of syntax: Commonalities and differences across children. *Dev. Sci.* 11, 84–97. doi: 10.1111/j.1467-7687.2007.00656.x

Wang, Y., and Tao, S. (2020). "The dynamic co-development of linguistic and discourse-semantic complexity in advanced L2 writing," in *Complex Dynamic Systems Theory and L2 Writing Development*, eds G. G. Fogal, and M. H. Verspoor (Amsterdam: John Benjamins Publishing Company), 49–78.

Wiese, H. (2020). "Language Situations: A method for capturing variation within speakers' repertoires," in *Methods in Dialectology XVI*, ed. Y. Asahi (Bern: Peter Lang), 105–117.

Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., et al. (2020). *RUEG Corpus*. doi: 10.5281/zenodo.3236068

Yager, L., Hellmold, N., Joo, H. A., Putnam, M. T., Rossi, E., Stafford, C., et al. (2015). New structural patterns in moribund grammar: Case marking in heritage German. *Front. Psychol.* 6:1–9. doi: 10.3389/fpsyg.2015.01716

Zimmer, C. (2020). Kasus im Namdeutschen. *Zeitschrift Für Germanistische Linguistik* 48, 298–335. doi: 10.1515/zgl-2020-2004

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fpsyg.2021.717615

Listeners' Linguistic Experience Affects the Degree of Perceived Nativeness of First Language Pronunciation

*Lisa Kornder and Ineke Mennen**

Department of English Studies, University of Graz, Graz, Austria

The aim of this study was to explore if and to what extent Austrian-English late sequential bilinguals who have been living in a second language (L2) environment for several decades are perceived to sound native in their first language (L1) when being compared to monolingual Austrian German (AG) control speakers. Furthermore, this investigation aimed to identify if listeners differ in their judgments of nativeness of L1 pronunciation depending on their own language background. For this purpose, two groups of native Austrian German listeners (N = 30 each), who differed regarding their linguistic background (Austrian German monolingual and Austrian German-English bilingual listeners) were asked to rate spontaneous speech samples produced by Austrian English bilingual and Austrian German monolingual speakers. Results showed that the bilingual L1 speech was perceived to sound overall less native compared to monolingual control speech. It was further observed that the two listener groups significantly differed in their perception of nativeness: Bilingual listeners were overall less likely to judge bilingual L1 pronunciation to sound non-native compared to monolingual listeners. To date, this is the first study to show that listener experience influences their perception of nativeness of L1 pronunciation and, thus, adds a new dimension to the notion of the native speaker.

Keywords: first language attrition, bilingualism, foreign accent, nativeness perception, English, (Austrian) German

OPEN ACCESS

Edited by:
Mila Vulchanova, Norwegian University of Science and Technology, Norway

Reviewed by:
Linda Wheeldon, University of Birmingham, United Kingdom
Roberto Ramírez Heredia, Texas A&M International University, United States

***Correspondence:**
Ineke Mennen
ineke.mennen@uni-graz.at

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

Received: *31 May 2021*
Accepted: *17 September 2021*
Published: *08 October 2021*

Citation:
Kornder L and Mennen I (2021) Listeners' Linguistic Experience Affects the Degree of Perceived Nativeness of First Language Pronunciation. Front. Psychol. 12:717615. doi: 10.3389/fpsyg.2021.717615

INTRODUCTION

A speaker's accent, shaped by various segmental and prosodic characteristics, is one of the most salient features of speech production and communication. Research shows that listeners are very sensitive to accented speech (e.g., Flege, 1984; Munro and Derwing, 1995b; Magen, 1998; Scales et al., 2006; Major, 2007; Eger and Reinisch, 2019) and that their perception and judgment of an individual is strongly influenced by the speaker's pronunciation (e.g., Lindemann, 2002; Kang and Rubin, 2009; Kang et al., 2016). In second language (L2) acquisition research, considerable attention has been given to the examination of features related to L2 learners' and bilinguals' accent, resulting from the observation that individuals who acquire an L2 relatively late in life often retain influences of their first language (L1) in their L2 pronunciation (Scovel, 1969; Flege, 1980, 1981; Flege et al., 1996; MacKay et al., 2001). This has led to acoustic investigations of the extent to which influences from the L1 lead to segmental and prosodic divergences from L2 production norms

(e.g., Flege, 1991; Thornburgh and Ryalls, 1998; Flege et al., 2003; Baker and Trofimovich, 2005; Simon, 2009; Levy and Law, 2010), as well as numerous studies that examine the perception of non-native L2 speech by (predominantly) monolingual listeners (e.g., Bongaerts et al., 1995; Flege et al., 1995; Munro and Derwing, 1995a; Riney and Flege, 1998; Moyer, 1999; Jilka, 2000; Kang et al., 2016).

While it is well-established that a speaker's L2 accent is likely to show traces of the L1 system, more recent research shows that the reverse is also possible, that is, a late-acquired L2 might affect a speaker's L1 accent due to bidirectional interaction processes taking place between the L1 and the L2 system (e.g., Mennen, 2004; Dmitrieva et al., 2010; Mayr et al., 2012; Bergmann et al., 2016; de Leeuw et al., 2017; Stoehr et al., 2017; de Leeuw, 2019). This phenomenon, referred to as *L1 attrition*, is frequently observed among bilinguals who have been long-term immersed in an L2 speaking country and who acquired their L2 after adolescence, i.e., at a point when the L1 is already fully developed in healthy individuals (Köpke and Schmid, 2004). Attrition research has provided evidence for the malleability of the L1 system with regard to segmental (e.g., de Leeuw et al., 2012a; Mayr et al., 2012; Stoehr et al., 2017; de Leeuw, 2019; Kornder and Mennen, 2021) and prosodic (de Leeuw et al., 2012b; Mennen and Chousi, 2018; de Leeuw, 2019; Gargiulo and Tronnier, 2020) features in L2-immersed late sequential bilinguals with different language backgrounds. These studies show that phonetic and phonological features of the L1 system might shift toward L2 production norms as a result of L2 learning experience and long-term exposure to the L2. The extent to which such L2-induced modifications in a bilingual's L1 accent are discernible for listeners has only recently started to attract attention (Sancier and Fowler, 1997; Hopp and Schmid, 2013; Schmid and Hopp, 2014; Bergmann et al., 2016; Mayr et al., 2020). Findings of accent rating studies show that listeners are indeed sensitive to divergences from L1 pronunciation patterns in bilingual speakers, which suggests that changes in L1 pronunciation resulting from long-term exposure to an L2 might lead to a detectable non-native L1 accent (e.g., De Leeuw et al., 2010; Bergmann et al., 2016; Mayr et al., 2020).

Following this line of inquiry, the present study aimed to contribute to the research on listener perceptions of potentially attrited L1 speech by conducting L1 nativeness ratings in a group of Austrian German-English late sequential bilinguals and monolingual Austrian German control speakers. In addition, this study set out to explore if and to what extent the linguistic background of listeners, that is, whether they are (quasi-)monolingual speakers of Austrian German or German-English bilingual speakers, affects their judgment of nativeness in L1 Austrian German pronunciation. In this context, we also examined if the extent to which bilinguals are perceived to sound (non-)native in their L1 pronunciation changes at different stages after immigrating to an L2 speaking country, judged by the rater groups described above. The decision to include two rater groups differing in terms of their language background is based on recent findings concerning L2 phonetic and phonological influences on L1 perception, showing that bilinguals' L1 perceptive abilities might be modified as a result of L2 learning experience (e.g., Celata and Cancila, 2010; Major, 2010; Carlson, 2018; Cabrelli et al., 2019). Hence, the present study sought to determine if differences in listeners' linguistic background may lead to differences in their perception of L1 nativeness.

First Language Phonetic Attrition

First language phonetic attrition refers to non-pathological modifications of L1 phonetic features resulting from long-term L2 learning experience and L2 exposure in bilinguals who acquired their L2 relatively late in life and who have been immersed in a migration setting for an extended period of time (e.g., Seliger and Vago, 1991; Köpke and Schmid, 2004). It should be noted though that L2-induced changes in the L1 system are not restricted to highly experienced bilinguals who have been permanently living in an L2 setting for a considerable period of time (see Kartushina et al., 2016a, for an overview). Bi-directional interaction processes between a speaker's linguistic systems leading to a shift of phonetic categories in the direction of the L2 might already occur at an early stage of L2 learning (e.g., Chang, 2012, 2013), as a result of recent and focused L2 production training (Dmitrieva et al., 2010; Kartushina et al., 2016b), or as a result of traveling between an L1 and an L2 country on a regular basis (Sancier and Fowler, 1997; Tobin et al., 2017). In these contexts, changes in the L1 pronunciation system are not indicative of L1 attrition, but represent instances of *phonetic drift* (Chang, 2012) of L1 categories toward L2 categories, that is, rather subtle L1 changes resulting from recent and increased L2 input. Phonetic drift, as described by Chang (2012, p. 264), can be considered as "one step in a continuum of cross-linguistic effects in bilinguals dependent on relative use of the L1 vs. the L2." Such subtle L2-induced modifications of the L1 pronunciation system occurring at an early stage of L2 learning have been shown to revert back to native L1 norms when speakers experience changes in their linguistic environment through, for instance, moving back to their L1 environment, and changes in language use (e.g., Chang, 2019). As such, instances of phonetic drift are not considered being indicative of L1 attrition given that they do not represent a decline in L1 proficiency (see e.g., Chang, 2012, 2013, 2019; Kartushina et al., 2016b). The present investigation focuses on potential L1 attriters, that is, late sequential bilinguals who are experienced L2 speakers and who have been living in an L2-speaking country for several decades, in which more persistent changes in L1 speech are reported to occur (e.g., Mayr et al., 2012; de Leeuw, 2019; Kornder and Mennen, 2021).

Empirical findings from research on L1 phonetic attrition provide evidence that a mature L1 pronunciation system is sensitive to L2 influences—in the same way a late-acquired L2 system is likely to be influenced by the native pronunciation system (e.g., Flege, 1980, 1981; Flege et al., 1996; Piske et al., 2001). These findings are in line with the Speech Learning Model (SLM) (Flege, 1995; Flege and Bohn, 2021), a model of L2 pronunciation development which predicts bidirectional influences between the L1 and L2 sound system, that is, the L1 influences the L2 system and vice versa. L1 segmental modifications in the direction of L2 norms have

been predominantly investigated with regard to bilinguals' productions of plosive consonants (Flege, 1987a; Mayr et al., 2012; Stoehr et al., 2017; Kornder and Mennen, 2021) and vowels (Mayr et al., 2012; Bergmann et al., 2016; de Leeuw, 2019; Kornder and Mennen, 2021). Other segmental features which have been shown to undergo attrition include, for instance, the post-vocalic rhotic consonant/r/ (Ulbrich and Ordin, 2014) and the lateral approximant/l/ (de Leeuw et al., 2012a; de Leeuw, 2019). By contrast, the extent to which L1 prosodic features are likely to be affected by L2-induced changes has been less frequently examined so far (de Leeuw et al., 2012b; Mennen and Chousi, 2018; de Leeuw, 2019), but findings suggest that also intonation features, such as tonal alignment (de Leeuw et al., 2012b; Mennen and Chousi, 2018), might shift in the direction of the L2.

The extent to which segmental and prosodic divergences from L1 norms lead to a perceived global foreign accent in L2-immersed late sequential bilinguals' L1 pronunciation has been explored in a series of foreign accent rating studies (FARs) (Sancier and Fowler, 1997; De Leeuw et al., 2010; Hopp and Schmid, 2013; Schmid and Hopp, 2014; Bergmann et al., 2016; Mayr et al., 2020). Overall, the findings of these studies confirm that L2-induced modifications in a speaker's L1 pronunciation system are in many cases distinctively discernible for listeners. De Leeuw et al. (2010), for example, examined the degree of perceived foreign accent in a group of German-English and German-Dutch bilinguals living in an L2 environment. Their findings revealed that monolingual German listeners were more likely to judge the L1 pronunciation of the bilingual group as sounding foreign compared to the speech of monolingual German controls. Similar observations were made by Bergmann et al. (2016) who found that nearly 40% of their German-English bilinguals received significantly higher foreign accent ratings compared to monolingual control speakers, that is, monolingual German listeners perceived some of the bilinguals to sound clearly non-native in their L1 German while others scored within the native range. The observation that not necessarily all bilinguals are perceived to have a non-native L1 accent was also made by Hopp and Schmid (2013), who showed that monolingual German listeners judged the majority of German-English and German-Dutch bilinguals (N = 29 out of 40) to sound not significantly differently from German control speakers in their L1 German pronunciation. At the same time, the ratings obtained for the remaining bilinguals (N = 11 out of 40) were clearly below the native German range.

While the accent rating studies outlined above focused on bilinguals with L1 German, Mayr et al. (2020) assessed perceived nativeness in monolingual Spanish speakers and two groups of native Spanish speakers of English. One group included Spanish-English bilinguals who taught Spanish as an L2 in the United Kingdom (teachers), the other group consisted of non-teachers with L1 Spanish who also lived in the United Kingdom and were late learners of L2 English. FAR results showed that monolingual Spanish listeners attributed overall higher accent scores to the L1 Spanish produced by the teachers compared to the non-teachers and monolingual Spanish controls. Mayr et al. (2020) argued that this difference might be explained by the fact that the teachers were immersed in professional environments with a high level of language co-activation and regular exposure to learners' non-native pronunciations. Similar observations were made by De Leeuw et al. (2010) who found that L1 accent ratings were overall highest for native German speakers of L2 English or L2 Dutch who were predominantly immersed in communicative settings in which code-mixing was enhanced, leading to higher levels of language co-activation. By contrast, monolingual German controls and German-English/Dutch bilinguals who were immersed in settings in which code-mixing was less likely to occur received comparatively lower accent ratings.

Another study frequently discussed in the context of perceived L1 accent was conducted by Sancier and Fowler (1997) who asked monolingual English and monolingual Brazilian-Portuguese listeners to rate speech samples produced by a Portuguese-English late bilingual. They aimed to identify potential changes in the speaker's L1 Portuguese and L2 English accent after staying in an L1 (Brazil) and L2 (United States) environment for several months at a time. While native English listeners did not perceive the speaker's L2 accent to be significantly different after staying in the United States and Brazil, respectively, native Brazilian-Portuguese listeners perceived the speaker's L1 pronunciation to be more accented after being exposed to English for an extended period of time. While this study is often cited in the context of L1 attrition research, it should be noted that the subject examined by Sancier and Fowler is—based on the definition of L1 phonetic attrition outlined above—not an attriter in the strictest sense given that she experienced regular changes in her linguistic environment. Nevertheless, the findings presented in this study do not only show that a bilingual's L1 accent is sensitive to recent and enhanced L2 input, but that the resulting pronunciation changes are also detectable for listeners.

Overall, the findings outlined above confirm that L2-induced phonetic and phonological modifications in a speaker's L1 system often lead to a non-native accent in the L1. Accent rating studies offer valuable insights into pronunciation differences between monolingual and bilingual speakers and show that listeners are sensitive to diverging L1 pronunciation patterns. Furthermore, they reveal that the extent to which L2-immersed bilinguals are perceived to have a foreign accent in their L1 varies, that is, bilingual speakers are not inevitably perceived to sound less native compared to monolingual controls (De Leeuw et al., 2010; Hopp and Schmid, 2013; Bergmann et al., 2016). There are several possible reasons for this: First, a bilingual's L1 pronunciation system is not necessarily affected by attrition, that is, some speakers will not show signs of phonetic attrition (e.g., Bergmann et al., 2016) or attrition in other linguistic areas. Hence, it is unlikely that we would perceive them as different from monolingual speakers. Second, not all acoustic changes a speaker's L1 system might undergo will be perceived as non-native or foreign by listeners. Research has shown that modification processes influence bilinguals' L1 system selectively rather than causing changes in the entire L1 system, which may result in both more pronounced changes on the one hand, and relatively subtle modifications on the other hand (e.g., Mayr et al., 2012; Bergmann et al., 2016; Kornder and Mennen, 2021). Some of these changes might be too subtle

to be perceived by listeners or—even if they are perceived—they may not necessarily be rated as non-native. Furthermore, research suggests that perceiving an individual's pronunciation as non-native is not based on a single acoustic-phonetic feature, but rather results from an interplay between and accumulation of different segmental and prosodic features which diverge from expected pronunciation patterns (e.g., Jilka, 2000; Mennen, 2004; Bergmann et al., 2016; Ulbrich and Mennen, 2016; van Maastricht et al., 2017). Therefore, even if acoustic-phonetic analyses reveal modifications in specific L1 segmental and/or prosodic features, they may only lead to a perceived non-native accent in combination with other phonetic and phonological changes (e.g., Bergmann et al., 2016).

The third reason for the observation that not all bilinguals are perceived to sound (non-)native to the same degree relates to listener-dependent variables which might affect their perception of accented speech, including their own language background, i.e., are they monolingual or bilingual speakers, and their familiarity with non-native or accented speech, i.e., are they exposed to non-native speech on a regular basis. The role of listener variables has predominantly been explored in the context of perceived accentedness of L2 speech (e.g., Thompson, 1991; Flege and Fletcher, 1992; Kennedy and Trofimovich, 2008; Eger and Reinisch, 2019), but recent research suggests that rater effects, such as their familiarity with the target language, may also affect the perception of L1 speech (Schmid and Hopp, 2014). There is to date no empirical research which set out to systematically examine how listeners' linguistic background and experience with an L2 influences their perception of L1 speech produced by monolingual and bilingual speakers—a gap the present investigation aims to fill, as will be further outlined below.

Listener Variables

As pointed out above, differences in foreign accent ratings are likely to result from both speaker-specific and listener-specific features, with the latter being predominantly investigated in L2 acquisition research. Thompson (1991), for instance, identified a significant relationship between listeners' linguistic experience and perceived nativeness of L2 English speech produced by L1 Russian speakers. Native English listeners were considered experienced if they were fluent in any additional language and had frequent contact with non-native English speakers. The L2 accent rating results showed that experienced listeners were less likely to rate English speech samples produced by L2 English subjects as non-native than inexperienced English raters who did not speak an L2 and who were not exposed to non-native English speech on a regular basis (see also Cunningham-Andersson, 1997). These findings corroborate Long's (1990) observation that listeners who have experience with other languages and live, for instance, in linguistically diverse communities are likely to be more lenient toward non-native accents compared to listeners who lack linguistic experience. Eger and Reinisch (2019) also showed a correlation between listener experience and foreign accent ratings, which, however, points in the opposite direction. They examined the extent to which native German learners' proficiency in L2 English affects their perception of German-accented English speech. Similar to the present study (see section "Speakers and Speech Samples," for more details), experience was defined as listeners' overall L2 proficiency, based on their self-reported skills in and frequency of speaking and listening in English. In Eger and Reinisch's (2019) study, listeners were asked to rate the "goodness" of the pronunciation of individual English target words produced by German learners of English. Their findings suggest that the more proficient listeners were in their L2, the less likely they were to accept German-accented productions as "good" representations of the L2 target, that is, an increase in L2 proficiency was correlated with a higher sensitivity to accented L2 speech.

Other studies examining the perception of L2-accented speech, by contrast, do not show a correlation between listeners' linguistic experience and accent ratings. Kennedy and Trofimovich (2008), for example, investigated the extent to which listeners' experience with Mandarin-accented English influences their perception of foreign accentedness, intelligibility and comprehensibility. While the findings showed that more experienced listeners perceived non-native English samples to be more intelligible, experience was not observed to have an impact on listeners' ratings of accentedness. Similarly, Flege and Fletcher (1992) did not find a significant relationship between linguistic experience, defined as listeners' familiarity with non-native speech, and perceived accentedness. Also Major (2007) found that listeners' familiarity with the language to be rated did not have a significant impact on their perception of (non-)native speech, that is, listeners were equally able to distinguish between native and non-native Brazilian-Portuguese speech samples.

The extent to which listeners' linguistic background and experience play a role in the perception of accented L1 speech has not been investigated in greater detail so far. The only study—to the best of our knowledge—which addresses the impact of listener-specific variables on the perception of not only non-native L2 speech but also on bilinguals' L1 pronunciation was conducted by Schmid and Hopp (2014) who examined how variation in FAR scores can be attributed to rater differences, among other variables. Their study involved three groups of speakers, namely monolingual German controls, L1 German speakers of L2 English or L2 Dutch (=L1 German attriters), and L1 English or L1 Dutch learners of German (=L2 German learners). Both native and non-native German raters who studied L2 English at university level in Germany were asked to rate German speech samples produced by the different speaker groups according to perceived foreign accent. Findings showed a correlation between raters' familiarity and contact with the target language German and foreign accent scores obtained for the monolingual German speaker group. That is, raters who were less familiar with German were more likely to judge monolingual German speech samples as non-native. No such correlations were demonstrated for the ratings obtained for the bilingual speech samples (i.e., L1 German attriters and L2 German learners), suggesting that listeners' familiarity and contact with a specific language only influence their perception of monolingual speech while their perception of potentially attrited and L2 speech remains largely unaffected. Based on these findings, Schmid and Hopp (2014, p. 383) conclude that "[v]ariation in raters can lead to shifts

in the absolute assessment of the strength of foreign accents on a given scale, with some raters apparently being more strict in the threshold of whom they judge to be native or native-like".

Objectives

As previously outlined, early linguistic research focusing on bilingualism and L2 acquisition was conducted under the premise that the L1 is protected against influences from a late-acquired L2 due to biological maturation processes (Lenneberg, 1967; Scovel, 1969). That is, a fully developed and mature L1 system was not considered to be modified or become less accessible in response to the acquisition of an additional language. Empirical investigations, however, questioned the stability of the L1 system in that they revealed that L2 learning experience can, in fact, lead to changes in a speaker's L1 system (e.g., Flege and Hillenbrand, 1984; Flege, 1987b; Major, 1992; Schmid, 2002; Mennen, 2004; Mayr et al., 2012; de Leeuw et al., 2017). From the perspective of a dynamic systems approach to language development, temporal or permanent modifications of a speaker's linguistic systems—including both the second and the L1—are regarded as an inherent part of language acquisition and development, resulting from dynamic, ongoing L1–L2 interactions (e.g., de Bot, 2007; de Bot et al., 2007). Hence, studies such as the present one contribute to arriving at a more profound understanding of the complexity and dynamics of developmental processes involved in language acquisition and of what might happen to a speaker's L1 pronunciation system when an additional language is acquired. The flexibility and malleability of a mature L1 system has attracted increasing attention among linguists and L2 researchers in the past decades, as the growing number of studies exploring the phenomenon of L1 attrition shows. These include empirical investigations of L1 attrition in all linguistic areas, such as syntax (e.g., Schmid, 2002; Tsimpli et al., 2004), the lexicon (e.g., Ammerlaan, 1996; Schmid and Jarvis, 2014), morphology (e.g., Altenberg, 1991), semantics (e.g., Jaspaert and Kroon, 1992; Pavlenko and Malt, 2011), and phonetics and phonology (e.g., Mayr et al., 2012; Bergmann et al., 2016; Cho and Lee, 2016; de Leeuw et al., 2017; Mennen and Chousi, 2018; de Leeuw, 2019). The development toward acknowledging that a bilingual's linguistic configuration is characterized by mutual L1–L2 interactions confirms Grosjean (1989, p. 13) former prediction that linguists "will no longer examine one of the bilingual's languages without examining the other," which gave rise to an integrated and holistic view on bilingualism (see also Grosjean, 1997). Moreover, the possibility that bilinguals might end up speaking not only their L2 but also their L1 with a detectable non-native accent is relevant from a sociolinguistic point of view, questioning the idealized image of the native speaker and pointing to the necessity to reconsider the native speaker norm and its relevance in linguistic research (see Davies, 2003). When it comes to the question of who is considered to be a native speaker and how nativeness is assessed, examining effects of listeners' language background and experience on their own perception of L1 speech is crucial. Not only do the findings of such investigations entail methodological consequences for the design of accent rating studies (see Schmid and Hopp, 2014, for a discussion), but they also add yet another dimension to the discussion of whether the native speaker concept in its traditional definition is still maintainable.

Based on the above considerations, the present study aimed to contribute to the emerging body of research exploring listener perceptions of potentially attrited L1 speech in order to determine if and to what extent L2-immersed bilinguals are perceived to have a non-native accent in their L1 when being compared to monolingual speakers (De Leeuw et al., 2010; Hopp and Schmid, 2013; Schmid and Hopp, 2014; Bergmann et al., 2016; Mayr et al., 2020). In addition, this study sought to investigate if listeners' language background and linguistic experience (monolingual vs. bilingual listeners) affect their perception of L1 pronunciation, that is, to find out whether there are any significant differences between monolingual and bilingual listeners in terms of their perception of nativeness in L1 speech. To this end, two groups of phonetically untrained listeners, (quasi-)monolingual Austrian German and bilingual AG-English listeners from Austria, were invited to rate a set of spontaneous L1 speech samples produced by AG-English late sequential bilinguals and monolingual AG controls according to perceived nativeness of pronunciation.

MATERIALS AND METHODS

Speakers and Speech Samples

Speech samples from AG-English bilinguals and AG monolinguals were extracted from publicly available German TV and radio interviews (see **Supplementary Table 1**). The bilingual group included three male AG-English celebrities, who are long-term United States immigrants and started acquiring English as an L2 in early adulthood. The first bilingual, Arnold Schwarzenegger (AS), was born in Thal, Styria (Austria), in 1947 and moved to the United States at the age of 21, where he made a career in bodybuilding, acting and politics (Schwarzenegger, 2012). Having learned English as a foreign language in an instructional setting in Austria, Schwarzenegger had only moderate English skills when he migrated to the United States (Outland Baker, 2006). Similar to Schwarzenegger, Frank Stronach (FS), who was born in a small municipality in East Styria (Austria) in 1932, left his home country in early adulthood and migrated to Canada in 1954 (Mayr, 2013; Noble, 2014). There, he started his first business which laid the foundation for a successful entrepreneur career. The third bilingual, Wolfgang Puck (WP), is an Austrian-born celebrity chef who moved from St. Veit an der Glan (Austria) to Los Angeles in 1973 where he started learning English (Schoenfeld, 2003).

The control group included five monolingual AG male speakers, aged 69–78, who were born and raised in Thal (Austria). Unlike the bilingual subjects, the control speakers were not well-known public figures, but locals who had been informally interviewed on different occasions in Thal. Given that the bilingual samples represented non-prompted, semi-spontaneous speech, the control samples were also selected from pre-recorded broadcast interviews, i.e., all speech samples included in the rating task were produced in a non-experimental setting. While the majority of accent rating studies rely on rehearsed

(e.g., Elliott, 1995; Derwing and Munro, 1997) or read (e.g., Thompson, 1991; Moyer, 1999) speech, this is, to our knowledge, the first L1 nativeness perception study which had listeners rate instances of non-prompted spontaneous speech. Using non-prompted speech samples can be considered being more representative of a speaker's natural and authentic pronunciation than speech samples elicited in a strictly controlled experimental setting (see e.g., Long, 1990).

From the broadcast interviews, bilingual and control samples were selected following a set of pre-defined criteria (see Jesney, 2004; Schmid and Hopp, 2014, for discussions). In order to ensure that listeners base their nativeness judgments on pronunciation-related features only, the samples did not include lexical and grammatical errors or hesitation and disfluency markers (e.g., Lennon, 1990). In addition, speech samples containing code-switches, high levels of background noise, longer pauses or self-corrections were excluded. Based on these criteria, a total of 28 speech samples was included in the speech corpus. For speaker AS, two different sets of samples were selected, one representing his early pronunciation in the late 1970s, and the other one representing his more current, late pronunciation in the 2010s. By comparing the ratings assigned to his early and late speech, we were able to explore if his late pronunciation was rated differently in terms of perceived nativeness compared to his early pronunciation. No such comparison could be drawn for the bilinguals FS and WP given that usable audio recordings representing their pronunciation at an earlier stage of migration were not available at the time the present study was conducted. Using speech samples produced by a single subject may constitute a limiting factor in our study, but the findings obtained in this single-subject investigation have the potential to serve as an incentive for future large-scale studies to explore if speakers' L1 pronunciation is likely to be perceived more or less native in the course of L2-immersion.

The individual samples varied in total duration, ranging from 1.95 to 5.39 s (M = 3.78, SD = 0.91; see **Table 1**). Previous L1 accent rating studies made use of considerably longer stretches of speech, in the range of approximately 10–20 s (e.g., Hopp and Schmid, 2013; Bergmann et al., 2016; Mayr et al., 2020). In the present study, however, shorter durations were selected to reduce the possibility that listeners identify the speakers, in particular the three bilinguals who, as mentioned above, are well-known celebrities. The length of the individual speech samples was considered sufficient for listeners to make their judgments given that we know from previous research that listeners are well able to rate speech samples which are very short in duration (see Flege, 1984). Empirical studies also show that listeners' ability to recognize speakers they are already familiar with is influenced by the duration of the speech samples they are presented with. Not surprisingly, speech samples which are longer in duration increase the likelihood that listeners identify a speaker (e.g., Schweinberger et al., 2013; see Mathias and von Kriegstein, 2014, for a discussion). However, research also indicates that listeners are well able to identify familiar speakers in speech samples which are shorter than 500 ms (see e.g., Fontaine et al., 2017). It has to be noted though that in these studies, listeners were specifically tested on their ability to recognize speakers based on voice recordings, i.e., listeners can be assumed to have made a special effort to identify the speaker, and were presented with a closed set of speakers to choose from. Recognition rates are reported to drop to above chance when identifying celebrities from an open response set (Van Lancker et al., 1984, 1985). In the present study, by contrast, the task was *per se* different in that listeners were not asked to pay attention to the speakers' identity, but rate their pronunciation according to perceived nativeness. After completing the rating task, listeners were asked on the rating sheet if they had noticed something about one or more of the speakers. Given that we did not want listeners to make a conscious effort to identify the speakers they were listening to, we did not specifically ask them *Did you recognize one or more of the speakers?*, but formulated a more open question, which still gave listeners the opportunity to mention if they had identified a speaker. In fact, N = 8 listeners, who were originally asked to complete the rating task, answered this question by stating that they had recognized speaker AS and/or speaker FS in some (but not all) of the speech samples. These listeners were excluded from analysis. In order to further ensure that listeners would not be able to detect who is speaking, the selected speech samples did not contain place and proper names which may uncover the identity of the speaker.

Listeners

A total of 60 listeners was recruited at the Department of English Studies at the University of Graz (Austria) and *via* personal contacts in Graz. Depending on their language background and English proficiency, the subjects were assigned to two different groups (N = 30 each). Monolingual (ML) raters (13 male, 17 female),[1] aged between 23 and 43 (M = 32.67, SD = 5.23), were (quasi-)monolingual[2] speakers of Austrian German, who reported having learned English in school for 5–9 years (M = 6.97, SD = 1.19), but rarely or never actively used English in private or professional contexts (see **Table 2**). Bilingual (BIL) raters (14 male, 16 female), aged between 22 and 27 (M = 24.1, SD = 1.32), were Austrian German learners of L2 English who were undergraduate students of English and American Studies at the University of Graz in their 2nd to 4th year (M = 3.05, SD = 0.67). Subjects in this group reported using English not only at university, but making moderate to frequent use of English in different communicative contexts outside university (see **Table 2**). Raters in both groups were born and raised in a monolingual AG environment in Graz and were permanent residents of Graz or surrounding areas. All participants reported normal hearing.

In order to obtain information concerning raters' linguistic background and language use, each subject was invited to fill in a questionnaire. Participants were asked to self-assess their overall English competence based on the six competence levels

[1]Raters were not matched according to age or gender given that there is no empirical evidence that these variables significantly affect accent ratings (see Schmid and Hopp, 2014).

[2]It is difficult—if not impossible—to find monolingual speakers in Austria who do not have at least some very basic knowledge of English (or other foreign languages) given that English has been a compulsory subject in Austrian schools since the mid–twentieth century (see de Cillia and Krumm, 2010).

TABLE 1 | Overview of speakers and speech samples.

	N subjects	N samples	Mean word count (SD)	Mean duration in sec. (SD)
Bilinguals	3	18	9.3 (2.06)	3.58 (0.88)
Controls	5	10	12 (3.27)	4.15 (0.89)
	N_{total} subjects	**N_{total} samples**	**Total word count**	**Total duration (sec.)**
	8	28	288	105.91

TABLE 2 | Listeners' self-reported frequency of English use in different contexts (1 = *Never*, 2 = *Very rarely*, 3 = *Rarely*, 4 = *Frequently*, and 5 = *Very frequently*).

	ML (*N* = 30)			BIL (*N* = 30)		
	Median	***Mean (SD)***	***Min-max***	***Median***	***Mean (SD)***	***Min-max***
Speaking (private)	*	*	*	3.5	3.5 (1.04)	1.0–5.0
Speaking (professional)	3.0	2.37 (0.99)	1.0–4.0	3.0	2.39 (1.38)	1.0–5.0
Watching (TV/films)	2.0	2.03 (1.07)	1.0–4.0	4.0	4.47 (0.51)	4.0–5.0
Reading	1.0	1.5 (0.73)	1.0–3.0	4.0	3.77 (0.68)	3.0–5.0
Listening	1.0	1.07 (0.25)	1.0–2.0	3.0	3.0 (1.34)	1.0–5.0
Writing	1.0	1.37 (0.72)	1.0–3.0	3.0	2.77 (1.45)	1.0–5.0

**All ML listeners reported that they never used English in private contexts. ML, monolingual listeners; BIL, bilingual listeners.*

defined by the Common European Framework of Reference for Languages (Council of Europe, 2001), which resulted in a six-point Likert scale ranging from 1 = very basic user to 6 = native or near-native user. ML listeners' self-assessed English proficiency ranged from 1 to 4 ($M = 2.6$, $SD = 0.89$) while the BIL listeners rated their proficiency in English from 4 to 6 ($M = 4.7$, $SD = 0.53$). Results of a Mann-Whitney U test showed that the two listener groups significantly differed in their self-assessed English proficiency ($U = 870$, $r = 0.829$, $p < 0.001$). In addition, each subject had to indicate if (*yes*/*no*) and how frequently (*very rarely*/*rarely*/*frequently*/*very frequently*) they used English across the four skills, i.e., speaking (professional vs. private), listening, reading, and writing. Their answers were converted to a five-point scale, ranging from 1 = never to 5 = very frequently. As shown in **Table 2**, participants in the BIL group used English overall more frequently in different contexts compared to the ML subjects, who used English either not at all or with low frequency only. Mann-Whitney U tests showed that the between-group differences in frequency of English use were significant for *Speaking (private)* ($U = 870$, $r = 0.877$, $p < 0.001$), *Watching* ($U = 876$, $r = 0.838$, $p < 0.001$), *Reading* ($U = 878$, $r = 0.844$, $p < 0.001$), *Listening* ($U = 814$, $r = 0.764$, $p < 0.001$), and *Writing* ($U = 699$, $r = 0.52$, $p < 0.001$). By contrast, the two groups did not differ in terms of frequency of English use in professional contexts ($p = 1$).[3]

Furthermore, the questionnaire collected information regarding subjects' familiarity with Austrian German varieties in order to avoid that they misperceive an Austrian German Styrian regional variety as a non-native accent (e.g., Flege et al., 1997). Participants in both groups reported being familiar with different Austrian German varieties, in particular with the Styrian variety spoken in Graz and surrounding areas.

Experimental Procedure

Due to the Covid-19-related restrictions in Austria, it was not possible to conduct the rating experiment in person. Therefore, rating materials and task descriptions were sent to the participants *via* e-mail. Each participant received a Power Point file, including the 28 test samples with one sound sample per slide, and detailed instructions. The rating sheet and the language background questionnaire were presented in an Excel file on two different spread sheets. As reported by some participants, it took approximately 15–20 min to complete the rating task.

The nativeness judgments were based on scalar ratings which have been used in previous accent rating studies to examine both perceived L2 (e.g., Moyer, 1999) and L1 (e.g., De Leeuw et al., 2010; Bergmann et al., 2016) accent. Raters were instructed to listen to one speech sample at a time and then (1) state whether the speaker is a native speaker of German (*yes*/*no*), and (2) indicate how certain they are concerning their judgment (*certain*/*relatively certain*/*uncertain*).[4] In the analysis process, listener answers were transferred to a six-point Likert scale, with 6 = certain of non-native speaker status, 5 = semi-certain of non-native speaker status, 4 = uncertain of non-native speaker status, 3 = uncertain of native speaker status, 2 = semi-certain of native speaker status, and 1 = certain of native speaker status. Resulting from this, speakers who received a low rating score were perceived to sound native or near-native in their German pronunciation while speakers with a high rating score were

[3]One reason for this might be that only some of the subjects included in the BIL group, but all ML subjects, had a (part-time) job given that they were university students at the time the present study was conducted.

[4]All instructions and questions were presented to the listeners in German. On the original rating questionnaire, each listener was asked (1) Hat dieser Sprecher Deutsch als Muttersprache? (Is this speaker a native German speaker?), and (2) Wie sicher sind Sie sich? (How certain are you of this?), with three answer options (sicher/certain; relativ sicher/relatively certain; and unsicher/uncertain).

perceived as non-native or near non-native speakers of German. The nativeness rating resulted in 1,680 individual rating scores (60 listeners × 28 ratings), which were averaged for the two rater groups (BIL vs. ML), for both speaker groups (bilinguals vs. controls) and for each of the bilinguals individually.

Statistical Analysis

In order to examine if the two speaker groups differed in the nativeness ratings they obtained and if the two listener groups differed in their perception of nativeness of bilingual and monolingual L1 speech, we ran a series of two-way repeated ordinal regression analyses in R (R Core Team, 2020), using the *Anova.clmm* function from the R package *RVAideMemoire* (Hervé, 2020) to determine main and interactions effects, including *rating score* as the dependent variable in each model. *Post hoc* Tukey's tests were conducted using the *emmeans* package (Lenth, 2020). An α-level of 0.05 was adopted throughout.

The first model (Model 1) was built to assess (1) if the two speaker groups (Bilingual vs. Control) differed in terms of perceived nativeness, and (2) if the two rater groups (BIL vs. ML) differed in terms of their perception of nativeness of monolingual and bilingual L1 speech. Model 1 included *rating score* as the dependent variable, *Rater_Group* (two levels: BIL, ML), *Speaker_Group* (two levels: Bilingual, Control) and an interaction between the two as independent variables, and *Rater*, *Speaker* and *Stimuli* as random factors. The model contained *Rater*-, *Speaker*-, and *Stimuli*-specific intercepts and by-*Rater* random slopes for *Speaker_Group*, by-*Speaker* random slopes for *Rater_Group*, and by-*Stimuli*-specific random slopes for *Rater_Group*. The second model (Model 2) aimed to test if the two listener groups (BIL vs. ML) differed in terms of their perception of nativeness of the individual bilingual speakers to determine if some of the bilinguals were perceived to be more or less native by the two listener groups, respectively. In this model, rating score was included as the dependent variable, and *Rater_Group* (two levels: BIL, ML), *Speaker* (nine levels: one for each of the speakers), and an interaction between *Rater_Group* and *Speaker* as independent variables. *Rater* and *Stimuli* were introduced as random factors, with random slopes for *Speaker* (by *Rater*) and *Rater_Group* (by *Stimuli*).

RESULTS

Interrater reliability was assessed for each of the two listener groups by calculating Cronbach's alpha coefficient. Inter-rater reliability was high in both the BIL (α = 0.84) and the ML (α = 0.89) listener group.

Figure 1 shows the rating scores obtained for the bilingual speaker group and the AG control group. On the six-point scale, ranging from 1 = *certain of native speaker status* to 6 = *certain of non-native speaker status*, monolingual control speakers received a median score of 1.0 (*min-max* = 1.0–5.0) and bilingual speakers were rated with a median score of 3.0 (*min-max* = 1.0–6.0), which suggests that the bilinguals were perceived to sound overall less native in their L1 compared to monolingual AG speakers. Model 1 (see section "Statistical Analysis") showed a main effect

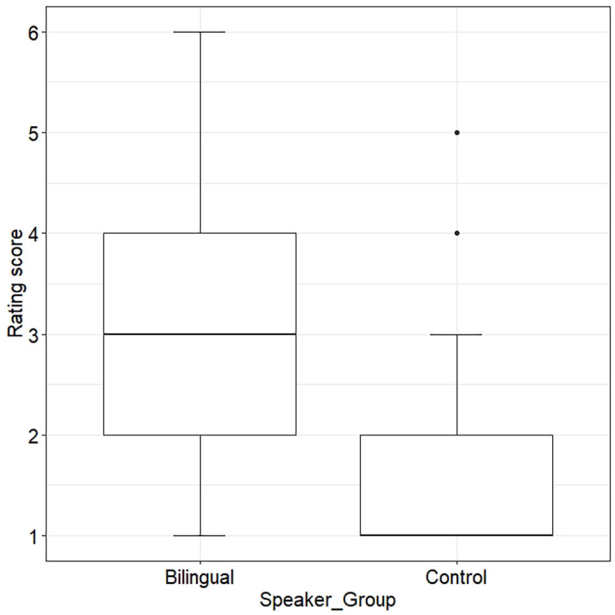

FIGURE 1 | Distribution of nativeness rating scores from 1 = *certain of native speaker status* to 6 = *certain of non-native speaker status* by speaker group (Bilingual vs. Control).

for *Speaker_Group*, $\chi^2[1] = 7.38$, $p < 0.001$, confirming that the two speaker groups significantly differed in terms of the rating scores they obtained, β = 4.77, *SE* = 0.866, $z = 5.51$, and $p < 0.001$.

Figure 2 displays the rating scores the two rater groups (BIL vs. ML) assigned to the bilingual and monolingual speaker group, respectively. ML listeners were overall more likely to rate the bilinguals' L1 pronunciation as sounding non-native (*Mdn* = 4.0, *min-max* = 1.0–6.0) compared to the BIL listeners (*Mdn* = 3.0, *min-max* = 1.0–5.0). At the same time, both rater groups judged the bilingual speakers to sound overall less native than the monolingual controls (*Mdn* = 1.0, min-max = 1.0–4.0). The statistical analysis (Model 1) revealed significant main effects for *Rater_Group*, $\chi^2[1] = 5.68$, $p < 0.001$, *Speaker_Group*, $\chi^2[1] = 7.38$, $p < 0.001$, and a significant interaction between the two, $\chi^2[1] = 17.75$, $p < 0.001$. Pairwise comparisons showed significant differences between the listener groups in terms of the nativeness ratings assigned to the bilingual speakers, β = −2.85, *SE* = 0.54, $z = -5.28$, and $p < 0.001$, that is, BIL listeners were less likely to rate bilingual speakers as non-native compared to ML listeners. By contrast, no significant differences between the rater groups were identified concerning their judgments of the control speakers ($p = 0.21$), i.e., both BIL and ML listeners perceived the monolingual controls to sound equally native.

Figure 3 shows the rating scores the two rater groups assigned to the individual bilingual speakers and the control group. The ML listener group rated the speech samples produced by AS_late with a median score of 4.5 (*min-max* = 1.0–6.0), and speakers FS and WP both received a median score of 4.0 (*min-max* = 2.0–6.0). By contrast, the three bilinguals received a considerably lower

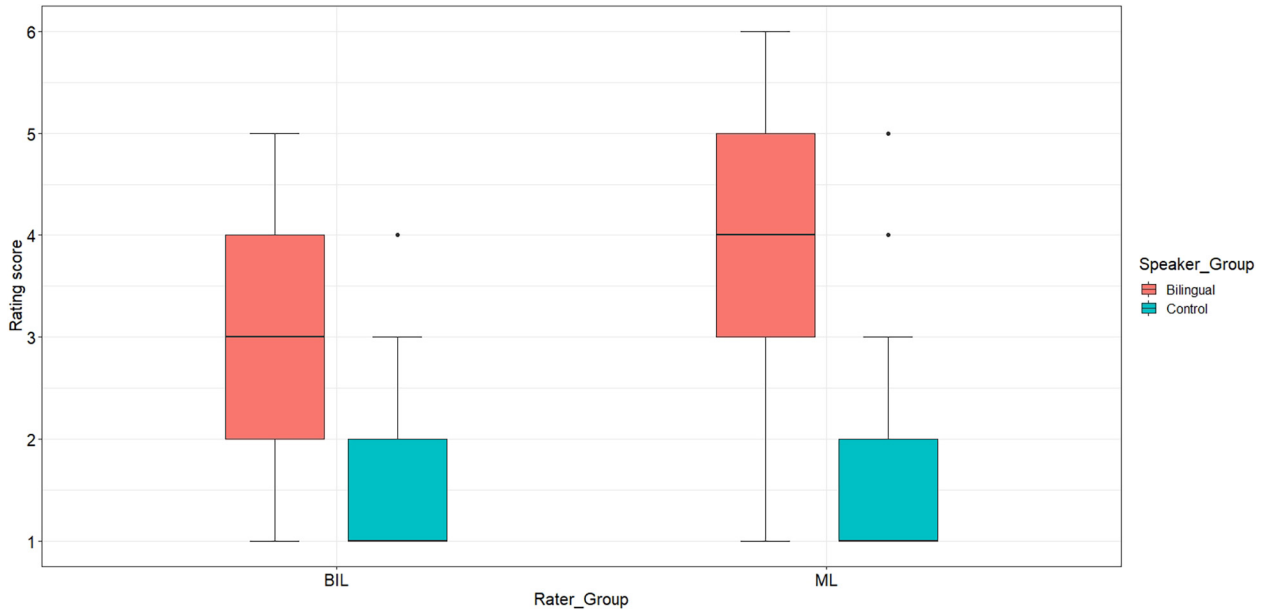

FIGURE 2 | Distribution of nativeness rating scores from 1 = *certain of native speaker status* to 6 = *certain of non-native speaker status* by rater group [Bilingual (BIL) raters vs. Monolingual (ML) raters].

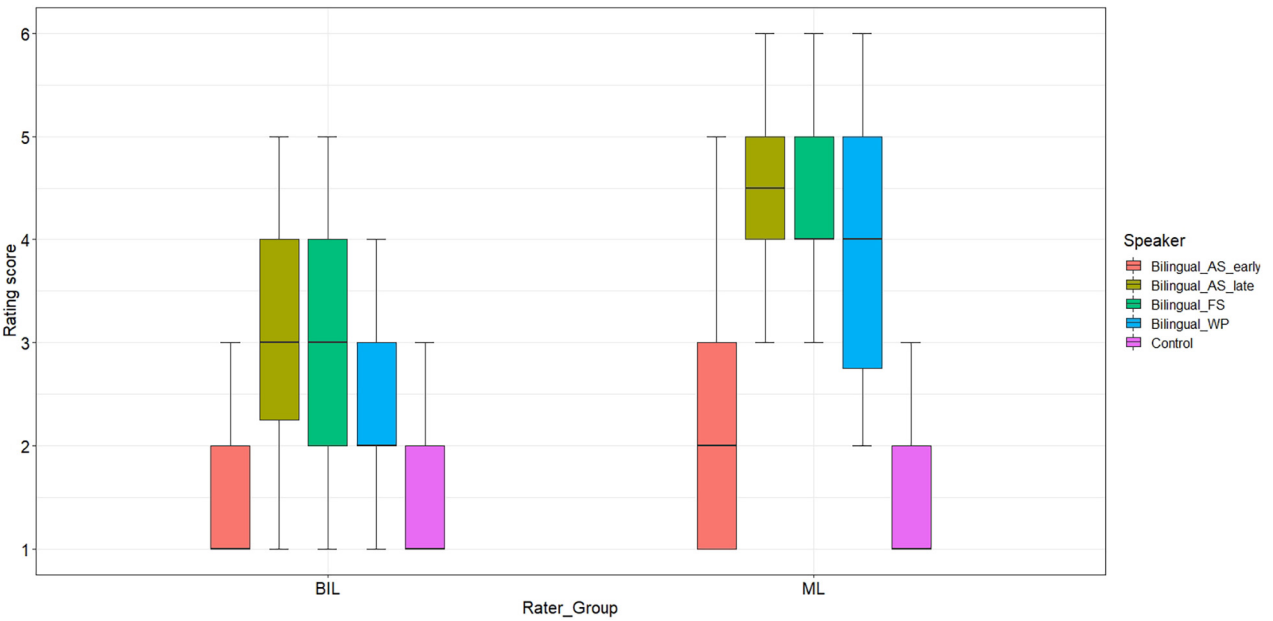

FIGURE 3 | Distribution of nativeness rating scores from 1 = *certain of native speaker status* to 6 = *certain of non-native speaker status* by rater group [Bilingual (BIL) raters vs. Monolingual (ML) raters] and individual speakers.

rating score from the BIL rater group (AS_late: *Mdn* = 3.0, *min-max* = 1.0–5.0; FS: *Mdn* = 3.0, *min-max* = 1.0–5.0; and WP: *Mdn* = 2.0, *min-max* = 1.0–4.0). Noticeably, both the BIL and the ML listeners judged speaker WP to sound more native compared to speakers AS_late and FS, but less native than the monolingual control group. It can be further observed that the speech samples produced by AS_early were rated to sound as native as the control speaker samples by the BIL listeners (*Mdn* = 1.0, *min-max* = 1.0–4.0) while the ML listeners judged AS_early to sound slightly less native (*Mdn* = 2.0, *min-max* = 1.0–5.0) than the monolingual controls (*Mdn* = 1.0, *min-max* = 1.0–5.0).

Model 2, including *Rater_Group*, *Speaker* and an interaction between the two as fixed factors, and *Rater* and *Stimuli* as random factors, showed significant main effects for *Rater_Group*,

$\chi^2[1] = 23.15$, $p < 0.001$, and *Speaker*, $\chi^2[8] = 405.88$, $p < 0.001$, as well as a significant interaction between *Rater_Group* and *Speaker*, $\chi^2[8] = 24.82$, $p < 0.001$. *Post hoc* pairwise comparisons revealed that the rating differences between the two listener groups were highly significant for AS_early, $\beta = -3.07$, $SE = 0.79$, $z = -3.87$, $p < 0.001$, AS_late, $\beta = -3.81$, $SE = 0.68$, $z = -5.6$, $p < 0.001$, FS, $\beta = -3.45$, $SE = 0.69$, $z = -5.07$, $p < 0.001$, and WP, $\beta = -3.21$, $SE = 0.76$, $z = -4.22$, $p < 0.001$, again confirming that BIL raters were less likely than ML raters to judge the bilinguals' L1 speech as non-native. While the rating differences between AS_late and FS were not shown to be significant in both the BIL ($p = 1.0$) and the ML ($p = 0.99$) rater group, the rating differences between WP and AS_late, BIL: $\beta = 1.83$, $SE = 0.46$, $z = 3.96$, and $p < 0.001$, ML: $\beta = 2.44$, $SE = 0.51$, $z = 12.13$, and $p < 0.001$, and WP and FS, BIL: $\beta = 1.91$, $SE = 0.44$, $z = 4.33$, $p < 0.001$, ML: $\beta = 2.2$, $SE = 0.4.7$, $z = 4.66$, and $p < 0.001$, turned out to be significant for both rater groups. That is, in both rater groups, speaker WP was perceived to sound more native in his L1 than AS_late and FS, but still less native than the monolingual control speakers, BIL: $\beta = -3.96$, $SE = 0.23$, $z = -17.2$, and $p < 0.001$; ML: $\beta = -5.72$, $SE = 0.25$, $z = -22.41$, and $p < 0.001$. The observation that the two rater groups differed in terms of their judgments of AS_early (see **Figure 3**) was also confirmed in the statistical analysis. ML listeners judged AS_early to sound less native in his L1 compared to the monolingual controls, $\beta = -2.59$, $SE = 0.27$, $z = -9.71$, and $p < 0.001$, while BIL listeners perceived his L1 pronunciation to be as native as the control speakers' AG pronunciation, that is, the rating differences between AS_early and the control group were not shown to be significant ($p = 0.26$). In addition, no significant differences were identified in terms of the ML and BIL listeners' rating scores assigned to the individual monolingual control speakers.

DISCUSSION

Perceived Nativeness of First Language Pronunciation

The present study aimed to (1) determine if long-term L2-immersed bilinguals are perceived to sound less native in their L1 when being compared to monolingual AG controls speakers, and (2) identify if and to what extent listeners' linguistic background (monolingual vs. bilingual) affects their perception of nativeness.

With regard to the first research aim, the nativeness rating experiment showed that, overall, bilingual speakers who started acquiring their L2 English as adults and who have been long-term immersed in an English-speaking country for several decades were, in fact, perceived to sound less native in their native AG pronunciation compared to monolingual AG control speakers. The observation that long-term L2 learning experience in an immersion setting may lead to a non-native L1 accent is in line with the findings of previous L1 accent rating studies (Sancier and Fowler, 1997; De Leeuw et al., 2010; Hopp and Schmid, 2013; Schmid and Hopp, 2014; Bergmann et al., 2016; Mayr et al., 2020). Taken together, these studies undermine the assumption that an individual's L1 system—once fully matured—remains impermeable and stable throughout the lifespan and is not prone to be altered by a late-acquired L2 system (e.g., Lado, 1957; Lenneberg, 1967; Scovel, 1969). Instead, the present and previous findings show that a bilingual's L1 system can indeed be modified resulting from a dynamic interaction between the two linguistic systems (see e.g., de Bot et al., 2007; de Bot and Larsen-Freeman, 2011; de Leeuw et al., 2012a), which might, in turn, entail that speakers' L1 pronunciation is perceived to sound less native when being compared to monolingual controls.

Alongside the observation that the L1 speech produced by bilingual speakers was perceived to sound considerably less native compared to monolingual speech, the present study revealed significant effects of listeners' linguistic background on their nativeness judgments. That is, native AG listeners who were actively learning L2 English and using their L2 in different contexts were more lenient in their judgments compared to (quasi-)monolingual AG listeners who were overall more likely to judge bilingual L1 speech to sound non-native. These findings add a further dimension to the observation that "[f]oreign accent is not only the way the learners produce the L2, but also what the native speakers of the target language *perceive* as such" (Reinisch, 2005, p. 82; our emphasis). As the present findings reveal, the same applies to a speaker's L1, which is not only influenced by speaker-related pronunciation features, but the extent to which a speaker is perceived to be a native speaker of their L1 also depends on who is listening. As addressed in Section "Listener Variables," previous studies examining potential effects of listener variables on the perception of L1 (Schmid and Hopp, 2014) and L2 (e.g., Thompson, 1991; Flege and Fletcher, 1992; Cunningham-Andersson, 1997; Kennedy and Trofimovich, 2008; Eger and Reinisch, 2019) speech provide rather inconclusive findings as to whether listeners' linguistic background and experience influence their perception of nativeness and accentedness. Some studies report an effect of listener experience on their judgments of nativeness, showing that linguistically experienced listeners are more tolerant toward non-native L2 speech compared to listeners without linguistic experience (Thompson, 1991). In the context of the present investigation, similar observations were made with regard to listeners' perception of L1 speech. Others, by contrast, have demonstrated that experienced listeners are more sensitive to accented speech and, therefore, are more likely to judge a speaker as sounding non-native (Eger and Reinisch, 2019), and still others have concluded that listener experience does not have a significant effect on nativeness judgments (Flege and Fletcher, 1992; Kennedy and Trofimovich, 2008). One reason for these—to some extent—contradictory findings might be that these studies use different criteria for classifying experienced vs. inexperienced listeners, ranging from self-reported L2 proficiency (e.g., Eger and Reinisch, 2019), being fluent L2 speakers with frequent non-native speaker contact (e.g., Thompson, 1991), to being familiar with non-native L2 speech, but not necessarily actively using an L2 (e.g., Flege and Fletcher, 1992; Major, 2007). It might be that being merely exposed to non-native speech without using an L2 in different communicative contexts is not sufficient to increase listeners' tolerance for accented speech. By contrast, being frequently exposed to non-native speech and using an L2 on a regular basis—as the listeners in Thompson (1991) and in the present study—makes listeners more lenient and perhaps less sensitive toward accented speech. In addition, it

must be taken into consideration that linguistic experience is, of course, not the only listener-dependent variable which might affect judgments of native vs. non-native speech. As pointed out by Munro et al. (2006), listener prejudices toward non-native accents may decrease their tolerance for accented speech while positive attitudes toward a particular accent are likely to make listeners being more lenient in their judgments of (non-)native speech (e.g., Beinhoff, 2013; Kraut and Wulff, 2013; Kraut, 2014). The extent to which additional listener variables, such as language attitudes, are correlated with listener experience and their perception of L1 pronunciation certainly offers a fruitful incentive for future research.

A third observation which has been made based on the results obtained in the nativeness rating study was that not all bilinguals were perceived to sound (non-)native to the same degree. For instance, both ML and BIL listeners perceived the bilingual speaker WP to sound more native in his L1 compared to the other two bilinguals, but still less native than the monolingual controls. The observation that bilingual speakers may differ in the degree of perceived nativeness—to the extent that some speakers are in fact not perceived differently from monolingual speakers (see section "First Language Phonetic Attrition")—has been made in previous L1 accent rating studies (e.g., Major, 1992; De Leeuw et al., 2010; Bergmann et al., 2016), which mirror the findings of studies focusing on perceived L2 accent (e.g., Bongaerts, 1999; Moyer, 1999). These studies show that even in groups of bilinguals who were carefully matched across a variety of factors, including age of L2 acquisition, proficiency, and length of residence in an L2 speaking country, speakers were not necessarily rated to sound equally (non-)native in their L2 speech. Similarly, some bilinguals might be perceived as relatively more or less native in their L1 pronunciation compared to other bilinguals. Alongside the reasons outlined in Section "First Language Phonetic Attrition," the observation that speakers may differ in the degree of perceived nativeness might account for an influence of additional internal and external factors, such as language learning aptitude or quantity and quality of target language input, which have been previously shown to affect the degree of non-native accent in L2 learning contexts (see Piske et al., 2001, for an overview). Only few investigations so far have systematically examined the role of such variables in the context of perceived nativeness of L1 speech (e.g., Hopp and Schmid, 2013). Hopp and Schmid (2013), for instance, found an effect of language aptitude on perceived foreign accent in the L1 speech of German-English and German-Dutch bilinguals, that is, speakers with high levels of language aptitude were, overall, judged to sound more native compared to speakers with comparatively lower levels of language aptitude, suggesting that language aptitude protects to some extent against L1 attrition effects. It should be noted though that Hopp and Schmid (2013) did not directly test language aptitude, but used a language proficiency test as an indirect measure of aptitude. Other factors, including frequency of language use and language attitudes, were not found to have a predictive effect on L1 nativeness ratings. Despite the fact that further systematic research is certainly necessary to get a better understanding of the role these factors effectively play when it comes to the perception of nativeness of L1 pronunciation and speakers' ability to retain a native accent in their L1, the observation made in the present study that not all bilinguals were perceived to sound equally non-native in their L1 speech might be explained against the background of additional influencing factors.

Interestingly, significant differences in the perception of nativeness were also observed when comparing the ratings obtained for the early (1970s) and late (2010s) speech produced by the bilingual AS. His early L1 pronunciation was judged to sound more native than his late pronunciation, but at the same time overall less native than the monolinguals' pronunciation—at least when considering the judgments made by the ML listeners. These findings show that the extent to which a speaker is perceived to sound native in their L1 might change over time in response to L2 learning experience and being immersed in an L2-setting. A previously conducted acoustic-phonetic investigation of AS's L1 vowels and plosives has, in fact, revealed a similar trend, namely that—in the course of L2-immersion—some of his L1 vowel and plosive targets have become less native-like, i.e., have moved away from native production norms (Kornder and Mennen, 2021). A reason why AS's early pronunciation was rated to sound less native compared to the monolinguals' AG pronunciation in the present investigation might be that the speech recordings representing his early pronunciation were made in the late 1970s, that is, approximately 10 years after he had migrated to the United States. Earlier recordings were not available at the time the present study was conducted. Hence, after a decade of L2 immersion, it can be assumed that AS had already gained quite some L2 learning experience which may have led to changes in his L1 pronunciation. As such, the present findings provide further evidence for the plasticity of a bilingual's L1 pronunciation system and show that a mature L1 is not robust over the lifespan. Given, however, that these observations were made for a single speaker only, future empirical studies need to examine such changes over time more closely, focusing on larger populations of bilingual speakers.

As outlined above, the observation that monolingual and bilingual listeners differ in their judgments of L1 pronunciation has been interpreted in the context of *listener tolerance*. We argue that speaking an additional language on a regular basis and use the L2 in different contexts increases listeners' tolerance for L1 speech which might diverge from expected L1 pronunciation norms. Conversely, listeners who lack frequent and regular exposure to an L2 seem to be more sensitive toward non-native pronunciation patterns, which makes them being less lenient in their nativeness judgments. However, an alternative explanation is that the use of an additional language has led to a perceptual restructuring of the L1. As mentioned in Section "First Language Phonetic Attrition," cross-linguistic interactions between a speaker's L1 and L2 linguistic system are not restricted to highly experienced and long-term immersed bilinguals, but have been observed to occur at different stages of bilingualism, that is, in both beginner L2 learners (e.g., Chang, 2012, 2013; Kartushina et al., 2016b) and highly experienced L2-immersed bilinguals (e.g., Mayr et al., 2012; Stoehr et al., 2017; de Leeuw, 2019; Kornder and Mennen, 2021). These L1–L2 interactions do not only lead to acoustic-phonetic changes in a bilingual's

L1 pronunciation system, but might also affect their ability to perceive a non-native accent in their L1 (see Major, 2010, for a discussion). In the context of the SLM; (Flege, 1995; Flege and Bohn, 2021), it is argued that a bilingual's L1 and L2 sound system mutually influence each other, which has been shown to lead to a restructuring of speakers' L1 phonetic categories in the direction of the L2 (e.g., Flege and Hillenbrand, 1984; Flege, 1987b, 1991; Mack, 1990). Moreover, the SLM predicts that "a strong bidirectional connection exists between production and perception" (Flege and Bohn, 2021, p. 29), that is, sound perception and production are considered to be interrelated. If we assume that L1-modifications resulting from L2 learning experience do not only influence speech production, but also affect speech perception, then it might be argued that the bilingual listeners in the present study have experienced underlying changes in their perception of L1 speech, triggered by their own bilingual background and L2 learning experience. Despite the fact that most studies examining phenomena of L1–L2 interactions in late sequential bilinguals focus on speech production either at the segmental level (e.g., Mayr et al., 2012; Stoehr et al., 2017; Kornder and Mennen, 2021) or at the level of global accent (e.g., De Leeuw et al., 2010; Bergmann et al., 2016; Mayr et al., 2020), some investigations set out to explore to what extent L1 speech perception in adult listeners is influenced by the L2 (Caramazza et al., 1973; Flege et al., 1999; Major, 2010; Alcorn and Smiljanic, 2017; Cabrelli Amaro, 2017; Carlson, 2018; Cabrelli et al., 2019). These studies either examine bilinguals' ability to discriminate native vs. non-native L1 pronunciation (Major, 2010), or assess bilingual listeners' perceptions of individual L1 segments, sound contrasts, or suprasegmental features (e.g., Caramazza et al., 1973; Flege et al., 1999; Alcorn and Smiljanic, 2017; Cabrelli Amaro, 2017; Carlson, 2018; Cabrelli et al., 2019). Some of these investigations provide evidence for a partial perceptual restructuring of the L1, showing that a late-acquired L2 does not only influence L1 production, but might also have an effect on L1 perception abilities (see e.g., Cabrelli et al., 2019, for Portuguese-English; Carlson, 2018, for Spanish-English; Celata and Cancila, 2010, for Lucchese-English). In the light of these findings, we might expect a restructuring of L1 perception also in the present study. That is, bilingual listeners' perception of L1 pronunciation might have been altered resulting from an interaction between their L1 and L2 system, which made them judge the nativeness of L1 pronunciation differently from listeners who do not speak an additional language. As a result, one and the same speaker might be perceived differently, depending on who is listening. This certainly entails methodological consequences for studies examining perceived global accent in that potential raters need to be carefully screened for their linguistic background and experience, acknowledging that these variables might influence their perception of L1 and presumably also L2 speech. Moreover, the observation that the extent to which a speaker is perceived to be a native speaker is influenced by a listener's own language background is relevant with regard to the concept of the native speaker, as will be further discussed below.

A last aspect which needs to be addressed when interpreting the findings obtained in the present study are potential range effects, which relate to the ratio of bilingual (or non-native) and monolingual speech samples represented in a rating task (see Flege and Fletcher, 1992; Schmid and Hopp, 2014). As Flege and Fletcher (1992) observed, listeners are more likely to judge non-native speakers as sounding foreign in rating tasks including a higher number of native control samples compared to non-native samples. In the present study, such range effects were reduced to some extent by including a higher number of bilingual samples (N = 18) and a comparatively lower number of monolingual samples (N = 10). Still, it should be taken into consideration that the bilingual samples were produced by three individual speakers only, which is—compared to previous accent rating studies (e.g., De Leeuw et al., 2010; Hopp and Schmid, 2013; Bergmann et al., 2016; Mayr et al., 2020)—a relatively small number. This may have influenced the reported results in that the three bilingual speakers might have stood out from the overall small number of speakers and were therefore more likely to be judged as sounding non-native in their L1 pronunciation.

Challenging "the Native Speaker"

Taken together, the findings of the present investigation do not only make an empirical contribution to the field of bilingual speech development in that they provide evidence for the malleability of the native language system and show that speakers might be perceived as non-native speakers of their L1, but they also add a new perspective to the broader notion of nativeness and the concept of the native speaker (e.g., Davies, 2003, 2004). Discussions related to the phenomenon of the native speaker have been evolving in various subfields of linguistics in the past decades, including, for example, sociolinguistics (e.g., Coulmas, 1981), language teaching (e.g., Leung et al., 1997; Cook, 1999), and multilingualism/bilingualism (e.g., Dostert, 2009; Rothman and Treffers-Daller, 2014). In its original, most basic definition, *the native speaker* has been described as a standard setter, portraying the embodiment of "true" language (see Davies, 2004), against which speakers are evaluated and judged. Evaluations and judgments of this kind do, in many cases, lead to accent-based discrimination, which so far has been predominantly considered in the context of non-native L2 speech (see e.g., Lippi-Green, 1997). But what happens if an individual is perceived to have a non-native accent in both their second and their first language? Despite the observation that there are non-native accents which listeners may consider more "acceptable" than others—including, for instance, prestigious European accents (Munro and Derwing, 2009)—speaking with a non-native accent in both the L1 and the L2 can entail serious personal and psycho-social consequences for individuals. These range from difficulties to access the job market (Munro, 2003), suffering from unequal payment (Dávila et al., 1993) to experiencing a profound strain on one's sense of belonging (Lippi-Green, 1997). Accent-based discrimination and stereotyping are inherently connected to the common belief that something like "true" nativeness—in the sense of accent-free speech—exists. However, as the findings of the present investigation illustrate, there is no single stable criterion based on which it would be possible to define *true* nativeness or the native speaker. If and to what extent speakers are perceived as native speakers depends on a variety of factors on behalf of both speakers and listeners. That is, the same speaker might

be perceived differently by different listeners, whose perceptions are shaped and influenced by their own linguistic backgrounds and presumably by additional variables, such as their attitudes toward specific languages and accents (e.g., Beinhoff, 2013; Kraut and Wulff, 2013; Kraut, 2014). Furthermore, the present findings reveal that a speaker might "lose" their status as native speaker over time, which further supports the view that the native speaker—in the sense of reflecting a stable and coherent concept—as such might not exist, particularly in the context of bilingualism. As pointed out by Cook (1999; see also Cook, 2003), bilinguals differ from monolingual speakers not only in their knowledge of an additional language, but also in terms of their cognitive processes, which essentially supports a holistic perspective on bilingualism (Grosjean, 1989). Hence, evaluating bilingual speech against monolingual standards and regarding monolingualism as the "benchmark of true nativeness," as Rothman and Treffers-Daller (2014, p. 93) put it, is rather misleading and does not properly reflect linguistic reality—considering that we live in a world where more than half of the population speaks more than one language (e.g., Aitchison, 1994; Grosjean, 2010). The observations made in the present study that bilingual speakers might no longer be perceived as native speakers of their L1 and that the perception of nativeness is strongly influenced by listeners' personal linguistic experience point to the need to reassess the static and idealized image of nativeness and to acknowledge the inherently dynamic nature of language and speech development.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Graz. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LK designed and conducted the experiment, analyzed and visualized the data, and wrote the manuscript. IM helped with the experiment design, wrote, reviewed, and edited the manuscript. Both authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Ulrich Reubold from the University of Graz for his useful statistical advice. A special thanks goes to the many participants who devoted their time to take part in the study. The authors acknowledge the financial support by the University of Graz.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717615/full#supplementary-material

REFERENCES

Aitchison, J. (1994). *Words in the Mind: An Introduction to the Mental Lexicon*, 2nd Edn, Cambridge, MA: Blackwell.

Alcorn, S. M., and Smiljanic, R. (2017). The production of epenthetic vowels by late Brazilian Portuguese English bilinguals. *Papers Presented at the Hispanic Linguistics Symposium*, Lubbock, TX.

Altenberg, E. P. (1991). "Assessing first language vulnerability to attrition," in *First Language Attrition*, eds H. W. Seliger and R. M. Vago (Cambridge: Cambridge University Press), 189–206. doi: 10.1017/CBO9780511620720.013

Ammerlaan, T. (1996). *'You Get a bit Wobbly…' Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition*. Ph.D thesis. Nijmegen: Katholieke Universiteit.

Baker, W., and Trofimovich, P. (2005). Interaction of Native- and Second-language vowel system(s) in Early and Late Bilinguals. *Lang. Speech* 48, 1–27. doi: 10.1177/00238309050480010101

Beinhoff, B. (2013). *Perceiving Identity Through Accent: Attitudes Towards Non-Native Speakers and Their Accents in English. Studies in Descriptive Linguistics*, Vol. 35. Oxford: Peter Lang. doi: 10.3726/978-3-0353-0454-1

Bergmann, C., Nota, A., Sprenger, S. A., and Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *J. Phon.* 58, 71–86. doi: 10.1016/j.wocn.2016.07.001

Bongaerts, T. (1999). "Ultimate attainment in L2 pronunciation: the case of very advanced late L2 learners," in *Second Language Acquisition and the Critical Period Hypothesis*, ed. D. Birdsong (Milton Park: Routledge), 133–159.

Bongaerts, T., Planken, B., and Schils, E. (1995). "Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis," in *The Age Factor in Second Language Acquisition*, eds D. Singleton and Z. Lengyel (Bristol: Multilingual Matters), 30–50.

Cabrelli, J., Luque, A., and Finestrat-Martínez, I. (2019). Infuence of L2 English phonotactics in L1 Brazilian-Portuguese illusory vowel perception. *J. Phon.* 73, 55–69. doi: 10.1016/j.wocn.2018.10.006

Cabrelli Amaro, J. (2017). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *Intern. J. Biling.* 21, 698–717. doi: 10.1177/1367006916637287

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., and Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French–English bilinguals. *J. Acoustic. Soc. Am.* 54, 421–428. doi: 10.1121/1.1913594

Carlson, M. T. (2018). Making room for second language phonotactics: effects of L2 learning and environment on first language speech perception. *Lang. Speech* 61, 598–614. doi: 10.1177/0023830918767208

Celata, C., and Cancila, J. (2010). Phonological attrition and the perception of geminate consonants in the Lucchese community of San Francisco (CA). *Intern. J. Biling.* 14, 185–209. doi: 10.1177/1367006910363058

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

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *J. Phon.* 41, 520–533. doi: 10.1016/j.wocn.2013.09.006

Chang, C. B. (2019). Language change and linguistic inquiry in a world of multicompetence: sustained phonetic drift and its implications for behavioral linguistic research. *J. Phon.* 74, 96–113. doi: 10.1016/j.wocn.2019.03.001

Cho, M.-H., and Lee, S. (2016). The impact of different L1 and L2 learning experience in the acquisition of L1 phonological processes. *Lang. Sci.* 56, 30–44. doi: 10.1016/j.langsci.2016.02.006

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Q.* 33, 185–209. doi: 10.2307/3587717

Cook, V. (2003). "The changing L1 in the L2 user's mind," in *Second Language Acquisition: Effects of the Second Language on the First*, Vol. 3, ed. V. Cook (Bristol: Multilingual Matters), 1–18. doi: 10.21832/9781853596346-003

Coulmas, F. (ed.) (1981). *A Festschrift for Native Speaker*. The Hague: Mouton de Gruyter. doi: 10.1515/9783110822878

Council of Europe (2001). *Common European Framework of Reference for Languages, Learning, Teaching, Assessment*. Available online at: https://www.europaeischer-referenzrahmen.de/sprachniveau.php (accessed March 8, 2020).

Cunningham-Andersson, U. (1997). "Native speaker reactions to non-native speech," in *Second-Language Speech: Structure and Process*, eds A. James and J. Leather (Berlin: Mouton de Gruyter), 133–144. doi: 10.1515/9783110882933.133

Davies, A. (2003). *The Native Speaker: Myth and Reality (Bilingual Education and Bilingualism)*, Vol. 38. Clevedon: Multilingual Matters, doi: 10.21832/9781853596247

Davies, A. (2004). "The native speaker in applied linguistics," in *The Handbook of Applied Linguistics*, eds A. Davies and C. Elder (Hoboken, NJ: Blackwell Publishing), 431–450. doi: 10.1111/b.9781405138093.2005.00023.x

Dávila, A., Bohara, A. K., and Saenz, R. (1993). Accent penalties and the earnings of Mexican Americans. *Soc. Sci. Q.* 74, 902–915.

de Bot, K. (2007). "Dynamic systems theory, lifespan development and language attrition," in *Language Attrition: Theoretical Perspectives*, Vol. 33, eds B. Köpke, M. S. Schmid, M. Keijzer, and S. C. Dostert (Amsterdam: John Benjamins Publishing Company), 53–68. doi: 10.1075/sibil.33.05bot

de Bot, K., and Larsen-Freeman, D. (2011). "Researching second language development from a dynamic systems theory perspective," in *A Dynamic Approach to Second Language Development: Methods and techniques*, eds M. Verspoor, K. de Bot, and W. Lowie (Amsterdam: John Benjamins), 5–24. doi: 10.1075/lllt.29.01deb

de Bot, K., Lowie, W., and Verspoor, M. (2007). A Dynamic systems theory approach to second language acquisition. *Biling. Lang. Cogn.* 10, 7–21. doi: 10.1017/S1366728906002732

de Cillia, R., and Krumm, H.-J. (2010). Fremdsprachenunterricht in österreich. *Sociolinguistica* 24, 153–169. doi: 10.1515/9783110223323.153

de Leeuw, E. (2019). Native speech plasticity in the German-English late bilingual Stefanie Graf: a longitudinal case study over four decades. *J. Phon.* 73, 24–39. doi: 10.1016/j.wocn.2018.12.002

de Leeuw, E., Mennen, I., and Scobbie, J. M. (2012a). Dynamic systems, maturational constraints, and L1 phonetic attrition. *Intern. J. Biling.* 17, 683–700. doi: 10.1177/1367006912454620

de Leeuw, E., Mennen, I., and Scobbie, J. M. (2012b). Singing a different tune in your native language: first language attrition of prosody. *Intern. J. Biling.* 16, 101–116. doi: 10.1017/S136672891100071X

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

de Leeuw, E., Tusha, A., and Schmid, M. S. (2017). Individual phonological attrition in Albanian-English late bilinguals. *Biling. Lang. Cogn.* 21, 278–295. doi: 10.1017/S1366728917000025

Derwing, T. M., and Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: evidence from four L1s. *Stud. Sec. Lang. Acquisit.* 20, 1–16. doi: 10.1017/S0272263197001010

Dmitrieva, O., Jongman, A., and Sereno, J. A. (2010). Phonological neutralization by native and non-native speakers: the case of Russian final devoicing. *J. Phon.* 38, 483–492. doi: 10.1016/j.wocn.2010.06.001

Dostert, S. C. (2009). *Multilingualism, L1 Attrition and the Concept of 'Native Speaker'*. Ph.D. thesis. Düsseldorf: University of Düsseldorf.

Eger, N. A., and Reinisch, E. (2019). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Stud. Sec. Lang. Acquisit.* 41, 179–200. doi: 10.1017/S0272263117000377

Elliott, R. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Mod. Lang. J.* 79, 356–371. doi: 10.1111/j.1540-4781.1995.tb01112.x

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Lang. Learn.* 30, 117–134. doi: 10.1111/j.1467-1770.1980.tb00154.x

Flege, J. E. (1981). The phonological basis of foreign accent: a hypothesis. *TESOL Q.* 15, 443–455. doi: 10.2307/3586485

Flege, J. E. (1984). The detection of French accent by American listeners. *J. Acoust. Soc. Am.* 76, 692–707. doi: 10.1121/1.391256

Flege, J. E. (1987a). "Effects of equivalence classification on the production of foreign language speech sounds," in *Sound Patterns in Second Language Acquisition*, eds A. James and J. Leather (Jacksonville: Foris), 9–39. doi: 10.1515/9783110878486-003

Flege, J. E. (1987b). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *J. Phonet.* 15, 47–65. doi: 10.1016/S0095-4470(19)30537-6

Flege, J. E. (1991). Age of learning affects the authenticity of voice–onset time (VOT) in stop consonants produced in a second language. *J. Acoust. Soc. Am.* 89, 395–411. doi: 10.1121/1.400473

Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Walmgate: York Press), 233–272.

Flege, J. E., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress,* ed. R. Wayland (Cambridge: Cambridge University Press), 3–83. doi: 10.1017/9781108886901.002

Flege, J. E., and Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *J. Acoust. Soc. Am.* 91, 370–389. doi: 10.1121/1.402780

Flege, J. E., Frieda, E. M., and Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *J. Phon.* 25, 169–186. doi: 10.1006/jpho.1996.0040

Flege, J. E., and Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *J. Acoust. Soc. Am.* 76, 708–721. doi: 10.1121/1.391257

Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *J. Acoust. Soc. Am.* 106, 2973–2987. doi: 10.1121/1.428116

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *J. Acoust. Soc. Am.* 97, 3125–3134. doi: 10.1121/1.413041

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (1996). "Factors affecting the production of word-initial consonants in a second language," in *Second Language Acquisition and Linguistic Variation*, eds R. Bayley and D. R. Preston (Amsterdam: John Benjamins), 47–74. doi: 10.1075/sibil.10.03fle

Flege, J. E., Schirru, C., and MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Commun.* 40, 467–491. doi: 10.1016/S0167-6393(02)00128-0

Fontaine, M., Love, S. A., and Latinus, M. (2017). Familiarity and voice representation: from acoustic-based representation to voice averages. *Front. Psychol.* 8:1180. doi: 10.3389/fpsyg.2017.01180

Gargiulo, C., and Tronnier, M. (2020). First language attrition on prosody in a foreign language environment: a speech production study on anaphora resolution. *J. Monoling. Biling. Speech* 2, 219–244. doi: 10.1558/jmbs.17093

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. (1997). The bilingual individual. *Interpreting* 2, 163–187. doi: 10.1075/intp.2.1-2.07gro

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674056459

Hervé, M. (2020). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics: R Package Version 0.9-77*. Available online at: https://CRAN.R-project.org/package=RVAideMemoire (accessed April 3, 2020).

Hopp, H., and Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: the impact of age of acquisition and bilingualism. *Appl. Psycholing.* 34, 361–394. doi: 10.1017/S0142716411000737

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: Benjamins), 137–147. doi: 10.1075/sibil.1.11jas

Jesney, K. (2004). The Use of Global Foreign Accent Rating in Studies of L2 Acquisition. A Report Prepared for the Language Research Centre University of Calgary. Available online at: https://www.semanticscholar.org/paper/The-Use-of-Global-Foreign-Accent-Rating-in-Studies-Jesney/3e4eaa880f359be90d48eb43df7574d98630926f (accessed September 27, 2021).

Jilka, M. (2000). "Testing the contribution of prosody to the perception of foreign accent," in *Proceedings of New Sounds (4th International Symposium on the Acquisition of Second Language Speech)*, Amsterdam, 199–207.

Kang, O., and Rubin, D. L. (2009). Reverse linguistic stereotyping: measuring the effect of listener expectations on speech evaluation. *J. Lang. Soc. Psychol.* 28, 441–456. doi: 10.1177/0261927X09341950

Kang, O., Vo, S. C. T., and Moran, M. K. (2016). Perceptual judgments of accented speech by listeners from different first language backgrounds. *Teach. Engl. Sec. Foreign Lang.* 20, 1–24.

Kartushina, N., Frauenfelder, U. H., and Golestani, N. (2016a). How and when does the second language influence the production of native speech sounds. *Lang. Learn.* 66, 155–186. doi: 10.1111/lang.12187

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2016b). Mutual influences between native and non-native vowels in production: evidence from short-term visual articulatory feedback training. *J. Phonet.* 57, 21–39. doi: 10.1016/j.wocn.2016.05.001

Kennedy, S., and Trofimovich, P. (2008). Intelligibility, comprehensibility, and Accentedness of L2 speech: the role of listener experience and semantic context. *Can. Mod. Lang. Rev.* 64, 459–483. doi: 10.3138/cmlr.64.3.459

Köpke, B., and Schmid, M. S. (2004). "Language attrition: the next phase," in *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, Vol. 28, eds M. S. Schmid, B. Köpke, M. Keijzer, and L. Weilemar (Amsterdam: John Benjamins), 1–43. doi: 10.1075/sibil.28

Kornder, L., and Mennen, I. (2021). Longitudinal developments in bilingual second language acquisition and first language attrition of speech: the case of Arnold Schwarzenegger. *Languages* 6, 1–25. doi: 10.3390/languages6020061

Kraut, R. (2014). Origins of language attitudes emerge in the perception of foreign-accented speech. *J. Sec. Mult. Lang. Acquisit.* 2, 30–38.

Kraut, R., and Wulff, S. (2013). Foreign-accented speech perception ratings: a multifactorial case study. *J. Multiling. Multicult. Dev.* 34, 249–263. doi: 10.1080/01434632.2013.767340

Lado, R. (1957). *Linguistics Across Cultures. Applied Linguistics for Teachers*. Ann Arbor, MI: University of Michigan Press, doi: 10.1017/S0008413100025196

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York, NY: John Wiley and Sons, doi: 10.1002/bs.3830130610

Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Lang. Learn.* 40, 387–417. doi: 10.1111/j.1467-1770.1990.tb00669.x

Lenth, R. (2020). *Emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.4.7*. Available online at: https://CRAN.R-project.org/package=emmeans (accessed April 3, 2020).

Leung, C., Harris, R., and Rampton, B. (1997). The idealised native speaker, reified ethnicities, and classroom realities. *TESOL Q.* 31, 543–560. doi: 10.2307/3587837

Levy, E. S., and Law, F. F. (2010). Production of French vowels by American-English learners of French: language experience, consonantal context, and the perception-production relationship. *J. Acoust. Soc. Am.* 128, 1290–1305. doi: 10.1121/1.3466879

Lindemann, S. (2002). Listening with an attitude: a model of native-speaker comprehension of non-native speakers in the United States. *Lang. Soc.* 31, 419–441. doi: 10.1017/S0047404502020286

Lippi-Green, R. (1997). *English With an Accent: Language, Ideology, and Discrimination in the United States*. New York, NY: Routledge.

Long, M. H. (1990). Maturational constraints on language development. *Stud. Sec. Lang. Acquisit.* 12, 251–285. doi: 10.1017/S0272263100009165

Mack, M. (1990). "Phonetic transfer in a French-english bilingual child," in *Plurilingua: IX. Language Attitudes and Language Conflict. Spracheinstellungen und Sprachkonflikte*, ed. P. H. Nelde (Berlin: Dümmler), 107–124.

MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (2001). Category restructuring during second-language speech acquisition. *J. Acoust. Soc. Am.* 110, 516–528. doi: 10.1121/1.1377287

Magen, H. S. (1998). The perception of foreign-accented speech. *J. Phonet.* 26, 381–400. doi: 10.1006/jpho.1998.0081

Major, R. C. (1992). Losing english as a first language. *Mod. Lang. J.* 76, 190–208. doi: 10.1111/j.1540-4781.1992.tb01100.x

Major, R. C. (2007). Identifying a foreign accent in an unfamiliar language. *Stud. Sec. Lang. Acquisit.* 29, 539–556. doi: 10.1017/S0272263107070428

Major, R. C. (2010). First language attrition in foreign accent perception. *Intern. J. Biling.* 14, 163–183. doi: 10.1177/1367006910363063

Mathias, S. R., and von Kriegstein, K. (2014). How do we recognise who is speaking? *Front. Biosci.* 6:92–109. doi: 10.2741/S417

Mayr, R., Price, S., and Mennen, I. (2012). First language attrition in the speech of Dutch–English bilinguals: the case of monozygotic twin sisters. *Biling. Lang. Cogn.* 15, 687–700.

Mayr, R., Sánchez, D., and Mennen, I. (2020). Does teaching your native language abroad increase L1 attrition of speech? The case of Spaniards in the United Kingdom. *Languages* 5:41. doi: 10.3390/languages5040041

Mayr, W. (2013). *Frank Stronach's Crusade to Transform Austria*. Available online at: https://www.spiegel.de/international/europe/frank-stronach-stirs-up-austrian-politics-with-new-party-a-885810.html (accessed February, 23, 2020).

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *J. Phon.* 32, 543–563. doi: 10.1016/j.wocn.2004.02.002

Mennen, I., and Chousi, D. (2018). "Prosody in first-generation adult immigrants and second-generation heritage-langauge users: the timing of prenuclear rising accents," in *Proceedings of the 9th International Conference on Speech Prosody*, New York, NY, 828–832. doi: 10.21437/SpeechProsody.2018-167

Moyer, A. (1999). Ultimate attainment in L2 phonology: the critical factors of age, motivation, and instruction. *Stud. Sec. Lang. Acquisit.* 21, 81–108. doi: 10.1017/S0272263199001035

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada J.* 20, 38–51. doi: 10.18806/tesl.v20i2.947

Munro, M. J., and Derwing, T. M. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Lang. Speech* 38, 289–306. doi: 10.1177/002383099503800305

Munro, M. J., and Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., and Derwing, T. M. (2009). Putting accent in its place: Rethinking obstacles in communication. *Lang. Teach.* 42, 476–490. doi: 10.1017/S026144480800551X

Munro, M. J., Derwing, T. M., and Morton, S. L. (2006). The Mutual Intelligibility of L2 Speech. *Stud. Sec. Lang. Acquisit.* 28, 111–131. doi: 10.1017/S0272263106060049

Noble, K. (2014). *Frank Stronach (Profile)*. Available online at: https://thecanadianencyclopedia.ca/en/article/frank-stronach-profile (accessed June 25, 2020).

Outland Baker, B. (2006). *Arnold and Me: In the Shadow of the Austrian Oak*. Bloomington, IN: Author House.

Pavlenko, A., and Malt, B. C. (2011). Kitchen Russian: cross-linguistic differences and first-language object naming by Russian-English bilinguals. *Biling. Lang. Cogn.* 14, 19–45. doi: 10.1017/S136672891000026X

Piske, T., MacKay, I. R. A., and Flege, J. E. (2001). Factors affecting the degree of foreign accent in an L2: a review. *J. Phon.* 29, 191–215. doi: 10.1006/jpho.2001.0134

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reinisch, E. (2005). What we (don't) perceive as foreign accent. *Grazer Linguist. Stud.* 64, 69–85.

Riney, T. J., and Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Stud. Sec. Lang. Acquisit.* 20, 213–243. doi: 10.1017/S0272263198002058

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

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

Scales, J., Wennerstrom, A., Richard, D., and Hui, W. S. (2006). Language learners' perceptions of accent. *TESOL Q.* 40, 715–738. doi: 10.2307/40264305

Schmid, M. S. (2002). *First Language Attrition, Use and Maintenance: The Case of German Jews in Anglophone Countries. Studies in Bilingualism*, Vol. 24. Amsterdam: John Benjamins. doi: 10.1075/sibil.24

Schmid, M. S., and Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: range and rater effects. *Lang. Test.* 31, 367–388. doi: 10.1177/0265532214526175

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

Schoenfeld, B. (2003). *Wolfgang Puck. Up Close and Personal With L.A.'s $375 Million Chef*. Available online at: https://www.winespectator.com/articles/wolfgang-puck-1785 (accessed April 5, 2021).

Schwarzenegger, A. (2012). *A Total Recall. My Unbelievably True Life Story*. London. New York: Simon & Schuster.

Schweinberger, S. R., Herholz, A., and Sommer, W. (2013). Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *J. Speech Lang. Hear. Res.* 40, 453–463. doi: 10.1044/jslhr.4002.453

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Lang. Learn.* 19, 245–253. doi: 10.1111/j.1467-1770.1969.tb00466.x

Seliger, H. W., and Vago, R. M. (eds) (1991). *First Language Attrition*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9780511620720

Simon, E. (2009). Acquiring a new second language contrast: an analysis of the English laryngeal system of natives peakers of Dutch. *Sec. Lang. Res.* 25, 377–408. doi: 10.1177/0267658309104580

Stoehr, A., Benders, T., van Hell, J. G., and Fikkert, P. (2017). Second language attainment and first language attrition: the case of VOT in immersed Dutch-German late bilinguals. *Sec. Lang. Res.* 33, 483–518. doi: 10.1177/0267658317704261

Thompson, I. (1991). Foreign accents revisited: the english pronunciation of russian immigrants. *Lang. Learn.* 41, 177–204. doi: 10.1111/j.1467-1770.1991.tb00683.x

Thornburgh, D. F., and Ryalls, J. H. (1998). Voice onset time in spanish-english bilinguals: early versus late learners of english. *J. Commun. Disord.* 31, 215–229. doi: 10.1016/S0021-9924(97)00053-1

Tobin, S. J., Nam, H., and Fowler, C. A. (2017). Phonetic drift in Spanish-English bilinguals: experiment and a self-organizing model. *J. Phon.* 65, 45–59. doi: 10.1016/j.wocn.2017.05.006

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. *Intern. J. Biling.* 8, 257–277. doi: 10.1177/13670069040080030601

Ulbrich, C., and Mennen, I. (2016). When prosody kicks in!. The intricate interplay between segments and prosody in the perception of foreign accent. *Intern. J. Biling.* 20, 522–549. doi: 10.1177/1367006915572383

Ulbrich, C., and Ordin, M. (2014). Can L2-english influence L1-German? The case of post-vocalic /r/. *J. Phon.* 45, 26–42. doi: 10.1016/j.wocn.2014.02.008

Van Lancker, D., Kreiman, J., and Emmorey, K. (1984). Recognition of famous voices forwards and backwards. *UCLA Work. Pap. Phon.* 59, 114–119. doi: 10.1121/1.2021007

Van Lancker, D., Kreiman, J., and Emmorey, K. (1985). Familiar voice recognition: patterns and parameters: part I: recognition of backwards voices. *J. Phon.* 13, 13–38. doi: 10.1016/S0095-4470(19)30723-5

van Maastricht, L., Zee, T., Krahmer, E., and Swerts, M. (2017). "L1 perceptions of L2 prosody: the interplay between intonation, rhythm, and speech rate and their contribution to accentedness and comprehensibility," in *Proceedings of Interspeech 2017*, Stockholm, 364–368. doi: 10.21437/Interspeech.2017-1150

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 October 2021
doi: 10.3389/fpsyg.2021.717449

Assignment of Grammatical Gender in Heritage Greek

*Demetris Karayiannis[1,2], Maria Kambanaros[2,3], Kleanthes K. Grohmann[2,4] and Artemis Alexiadou[5,6]**

[1] European Master's in Clinical Linguistics, Faculty of Arts, University of Groningen, Groningen, Netherlands, [2] Cyprus Acquisition Team (CAT Lab), University of Cyprus, Nicosia, Cyprus, [3] UniSA Allied Health & Human Performance, University of South Australia, Adelaide, SA, Australia, [4] Department of English Studies, University of Cyprus, Nicosia, Cyprus, [5] Institute of English and American Studies, Humboldt-Universität zu Berlin, Berlin, Germany, [6] Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany

This study investigates the acquisition of grammatical gender in Heritage Greek as acquired by children (6–8 years of age) and adolescents (15–18 years) growing up in Adelaide, South Australia. The determiner elicitation task from Varlokosta (2005) was employed to assess the role of morphological and semantic cues when it comes to gender assignment for real and novel nouns. Ralli's (1994) inflectional classes for Greek nouns and Anastasiadi-Symeonidi and Cheila-Markopoulou's (2003) categories of prototypicality were employed in the analysis of the collected data. The performance of heritage speakers was compared to that of monolingual speakers from Greece (Varlokosta, 2011). The results indicate that–beyond age differences in the two groups–a formal phonological rule guides gender assignment in the production of heritage speakers which departs from initial expectations.

Keywords: grammatical gender, heritage grammars, Modern Greek, bilingual language acquisition, inflectional class

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez, University of the Balearic Islands, Spain

Reviewed by:
Ji Young Kim, University of California, Los Angeles, United States
Yvonne Van Baal, Norwegian University of Science and Technology, Norway

***Correspondence:**
Artemis Alexiadou
artemis.alexiadou@hu-berlin.de

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

Received: *31 May 2021*
Accepted: *31 August 2021*
Published: *08 October 2021*

Citation:
Karayiannis D, Kambanaros M, Grohmann KK and Alexiadou A (2021) Assignment of Grammatical Gender in Heritage Greek. Front. Psychol. 12:717449. doi: 10.3389/fpsyg.2021.717449

INTRODUCTION

Gender has been argued to be one of the most elusive features of noun phrases (Corbett, 1991). According to Corbett (1991), languages fall into three types when it comes to their gender systems: There are languages with semantic gender systems, languages with predominantly semantic, mixed gender systems, and, finally, languages with formal gender systems. The first two groups use semantic criteria to distinguish between nouns (biological gender), while in the third group of languages, every noun carries gender specification and thus is classified in a particular way. Greek belongs to this group of languages, having a three-gender system, while English has a semantic gender system. Despite its role in nominal classification in this group of languages, as also observed by Picallo (2017), it is not completely straightforward what the role of gender is in a minimalist theory of grammar (Chomsky, 1995 and subsequent work). It is very different from all other types of features, as it does not seem to have a computational import. However, and in spite of the lack of computational import, in languages such as Greek gender seems to be acquired early.

Tsimpli and Hulk (2013) argue that language learners follow two stages in the acquisition of grammatical gender. First, they discover that their language has grammatical gender. Second, they become aware of the fact that gender is a feature to classify nouns in their language and can thus form generalizations and acquire the gender of individual nouns–but importantly also make predictions for novel nouns that they may encounter. While in some languages, such as Dutch, learners have difficulties identifying gender, the acquisition of gender in Greek is effortless, and thus learners acquire the gender system very early on (Tsimpli and Hulk, 2013;

Unsworth et al., 2014). For instance, Stephany (1997) reports that gender is acquired by the age of 2:3 (years:months); see also Mastropavlou (2006), who reports a target-like gender system for Greek at the age of 3:6. Neuter is considered to be the default gender for inanimate nouns and in case no agreement can be established (cf. Kazana, 2011; Mastropavlou and Tsimpli, 2011; Anagnostopoulou, 2017; Alexiadou et al., 2021).

Tsimpli and Hulk (2013: 128–129) introduce a further distinction: Next to the learner default (i.e., "the setting adopted by the learner in the earliest stage when input is either unavailable or unanalyzed as yet"), there is the notion of the linguistic default. This notion of a default has also been used to apply to a form which is based on the "elsewhere condition." From this perspective, it is the less specified value (which applies "elsewhere") of a particular feature that constitutes the linguistic default. As they argue, neuter is both the learner and the linguistic default in Greek, and this is the reason why its acquisition takes place at a very early age. While Tsimpli and Hulk (2013) did not differentiate between animate and inanimate nouns, Kazana (2011) and Anagnostopoulou (2017) argue that the default is regulated by animacy: In gender resolution contexts involving coordinated nouns, the two groups behave differently; in the case of animate nouns, conjoined singular and plural nouns of the same gender resolve to the gender of the conjuncts. By contrast, conjoined nouns of mixed gender resolve to masculine or feminine if the conjoined nouns denote females. In the case of inanimate nouns, the situation is different: Nouns with the same gender resolve to the gender value of the conjuncts, while inanimate nouns with mixed gender resolve to neuter. This suggests that there is a difference between the learner default and the linguistic default in the case of animates–while the learner default is neuter, the linguistic default is masculine.

This paper reports a pilot study which approaches the notion of the "native speaker" from the heritage language perspective, by focusing on the acquisition of grammatical gender in Heritage Greek in South Australia. The goals of this study were on the one hand to establish whether a determiner elicitation task is a suitable tool for the population in question, and on the other hand to provide novel data on the accuracy of heritage Greek speakers in assigning grammatical gender to real nouns, to investigate whether the lack of semantic information is a significant factor in gender assignment in novel nouns, and to investigate to what extent the same lexical, morphological, and referential principles guide the grammatical gender assignment in the mental grammar of heritage and homeland speakers. After providing the relevant background on heritage linguistics, gender in heritage languages with emphasis on Greek, and the nominal system of Greek, we will present the research (methodology and results), discuss the findings, and briefly conclude.

BACKGROUND

Heritage Linguistics

The role of cross-linguistic influence has been discussed in detail in studies of bilingual language development. Focusing on languages that have gender systems, the main issue has been as to whether or not this would lead to an acceleration or a delay of acquisition. In heritage linguistics, Polinsky (2008) discusses the re-organization of the Russian gender system from a three-gender language to a two-gender language in correlation with the level of proficiency, while Laleko (2019) reports that Russian heritage speakers encounter greater difficulties with underspecified forms, that is, forms that could belong to more than one gender, for which they rely on morpho-phonological criteria.

The discussion of gender in heritage Russian is interesting in the context of Greek, as both languages use a variety of criteria to assign gender: (a) lexical, where certain forms are inherently specified for gender (e.g., *pateras* "father"); (b) morpho-phonological, where certain nouns bear a particular feminine suffix (e.g., *furnar–is–a* "female baker"); and (c) referential, where certain nouns are not specified for gender but receive gender specification via association with a human referent in discourse (see e.g., Alexiadou, 2004, 2017).

Most of the literature on heritage languages focuses on gender agreement. Thus, Polinsky (2018: 206) observes that gender agreement shows effects of vulnerability in heritage speech independently of the gendered vs. un-gendered nature of the language these speakers are dominant in. Montrul et al. (2008: 515) state that "gender agreement appears to be a strong candidate for language loss in a language contact situation."

There are three studies on heritage Greek which also primarily look at gender agreement. Paspali (2019) tested gender agreement with adult heritage speakers of Greek in Germany, the result being that her speakers were not statistically different from the monolingual controls. Kaltsa et al. (2017) deal with the acquisition of gender assignment and agreement with Greek–English and Greek–German bilinguals. For the gender agreement tasks, their findings suggest that "neuter and masculine are discriminated since the former shows significantly better scores compared to masculine and feminine shows the lowest performance" (p. 24). In addition, they note "that both bilingual groups performed similarly in the gender agreement tasks and were significantly more accurate with neuter than with masculine and feminine suggesting that neuter is treated as the default value, giving rise to fewer errors" (p. 24). Alexiadou et al. (2021) focused on adolescent and adult heritage speakers of Greek in the United States and found that the participants show mismatches in gender agreement, which differ with respect to the agreement target between groups. In particular, noun phrase-internal agreement seems more affected in the adolescent group, while personal pronouns appear equally affected. Their results suggest that heritage speakers resort to neuter gender, which can be attributed to the fact that they have difficulties with establishing agreement chains.

In order to substantiate the use of a default strategy, unlike these previous studies we focus on gender *assignment*, which we investigate in two different age groups.

The Greek Nominal System

Modern Greek is a highly inflected language. In its nominal system, it syncretically encodes the phi-features of number (singular and plural, henceforth Sg/Pl), case (nominative, accusative, genitive, and vocative, henceforth

TABLE 1 | Overview of the Greek inflectional classes according to Ralli (1994).

IC1 (−*os* M/F)		**IC2 (−*s* M)**			**IC3 (−Ø F common)**		**IC4 (−Ø F learned)**
o aðelf-os (M) "brother"	i isoð-os (F) "entrace"	o pater-a-s "father"	o kaf-e-s "coffee"	o xoreft-i-s "dancer"	i miter-a-Ø "mother"	i kor-i-Ø "daughter"	i poli-Ø "city"
IC5 (−*o* N)		**IC6 (−*i* N)**			**IC7 (−*os* N)**		**IC8 (−*ma* N)**
to ner-o "water"		to xtapoð-i "octapus"			to var-os "weight"		to ci-ma "wave"

The determiner precedes the noun.

Nom/Acc/Gen/Voc), and grammatical gender (masculine, feminine, and neuter, henceforth M/F/N). Nouns of native stock and integrated loanwords overwhelmingly fall into eight morphological paradigms, although the number of nouns belonging to each class is not equally distributed. The paradigms are determined by shared morphology across Case and Number (Ralli, 1994, 2002, 2003). These eight inflection classes, henceforth ICs 1–8, are presented in **Table 1** (see also Alexiadou and Müller, 2008).

In six out of the eight inflectional classes, the gender-marking suffix in Nom.Sg is realized a consonant (e.g., *−s* for IC2), a vowel (e.g., *−i* for IC6), or a syllable (e.g., *−ma* for IC8). In the case of IC3 and IC4, the presence of a null suffix is proposed in order to elegantly capture the observation of a uniform inflectional paradigm between the feminine nouns with phonological forms ending in/i/and/a/(those vowels being described as parts of the lexical root, or *thematic*), such as *miter-a-Ø*$_{\text{NOM}}$ "mother"; *kor-i-Ø*$_{\text{NOM}}$ "daughter" and the respective forms *miter-a-s*$_{\text{GEN}}$; *kor-i-s*$_{\text{GEN}}$ (Ralli, 1994, 2002). The two feminine noun categories are maintained as distinct due to their different plural paradigm in all cases except the genitive: *−es*$_{PL}$ for IC3, but *−is*$_{PL}$ for IC4. The words in IC4 belong to a more heightened language register but form a productive class. The members of IC8 are deverbal nouns ending in *−ma*. They show imparisyllabic inflected forms with an additional syllable (e.g., *ci.ma*$_{\text{NOM}}$; *ci.ma.tos*$_{\text{GEN}}$, "wave"). As such they are distinct from some feminine IC4 stems ending in/m/followed by a thematic/a/ (cf. mam-a-Ø, "mom").

A small number of native words exhibit idiosyncratic declension patterns and do not neatly fit under any IC; moreover, most recent loanwords have no overt inflection (but their phi-features are still valued, as revealed by agreeing determiners and adjectives). The task includes nouns from all eight regular inflectional classes as well as two native lexical items with irregular declension which will subsequently be referred to as "imparisyllabic neuter *−s*," due to their different syllable count in the genitive singular and all plural forms. These nouns are part of a small non-productive class of neuter nouns bearing the suffix *−s* whose genitive and plural forms are reminiscent of those of IC8 nouns (e.g., *kre.as*$_{\text{NOM}}$; *kre.a.tos*$_{\text{GEN}}$, "meat").

The determiner system reflects the same phi-features. They precede and agree with the noun for gender, case, and number. In this study, the nominative singular forms of the definite determiner were elicited: o$_{\text{M}}$, i$_{\text{F}}$, and to$_{\text{N}}$.

The gender system of Greek is described as a primarily formal system, where morphological form, rather than semantic features, predicts the gender value of a given word (Varlokosta and Nerantzini, 2013). This is in opposition to primarily semantic systems, or mixed systems, where the biological sex of the referent is the main predictor of a noun's grammatical gender, such as in English. The systems can also differ in the extent to which gender is marked across different categories: Greek gender is marked across different nominal and determinative categories, English gender is restricted to the pronominal category.

The role of real-word sex information is elevated under a different categorization system proposed by Anastasiadi-Symeonidi and Cheila-Markopoulou (2003), which relies on the notion of "prototypicality." According to these authors, certain nouns are prototypically masculine, for example: These are animate, their referent is human, and they bear the morphological ending *−s* (e.g., *pateras* "father"). Other masculine nouns are non-prototypical, and they are inanimate (*çimonas* "winter"). Similarly, prototypical female nouns are animate, their referent is female, and they end in *−a*, *−i*, and *−u*. Finally, inanimate neuter nouns are prototypical and they end in *−o*, *−i*, and *−a*. By contrast, non-prototypical neuter nouns include inanimate nouns ending in *−s* and *−n* and animate nouns (for animals) both inflected (*provato* "sheep") and uninflected (*koala*) as well as uninflected human nouns (*barman* "barman").

We note that the suffixes employed by this framework do not line up with those proposed by Ralli (1994), who teases apart gender suffixes that coincide in the citation form but differ during inflection (e.g., *mam–a*$_{\text{F.NOM}}$; *mam–as*$_{\text{F.GEN}}$ vs. ci-ma$_{\text{N.NOM}}$; *cima–tos*$_{\text{N.GEN}}$). Instead, we adopt Anastasiadi-Symeonidi and Cheila-Markopoulou's framework in accordance with the design of determiner elicitation task from Varlokosta (2005).

By employing those two systems of categorization, namely Ralli's exclusively morphological categorization and Anastasiadi-Symeonidi & Cheila-Markopoulou's system that incorporates semantic as well as morphological cues, we are able to evaluate the contribution of both types of cues separately in the real noun condition.

METHODOLOGY

For the purposes of this pilot study, the accuracy of grammatical gender assignment was measured by eliciting the appropriate definite article form in a determiner–noun context. Varlokosta's (2005) "Test for Grammatical Gender Assignment to Greek Nouns" was employed, a task which consists of real nouns coming from all ICs as well as novel nouns ending in the seven different possible suffixes for Greek nouns. The participating children were

asked to produce the corresponding singular nominative form of the article following a short practice round.

Participants

A total of 37 young heritage speakers of Greek were recruited in Adelaide, South Australia. The participants were raised and, at the time of testing, resided in Australia. All of them attended English-medium education in Adelaide and as such we consider their dominant language to be English. They were classified in two distinct age groups: 24 children aged 6–8 years (M = 8.01, SD = 0.64) and 13 adolescents aged 15–18 years (M = 16.77, SD = 1.03). The population in question and the sample of the current pilot study is small and heterogeneous, owing to both the inherent characteristics of the Heritage Greek community of Australia, and the material and time constraints faced by the researchers. No additional factors were considered–such as socio-economic family background, country of birth, age of onset of acquisition for the dominant and heritage languages, formal schooling or literacy in the heritage language, or knowledge of additional languages beyond Greek and English–as priority was given to establishing the appropriateness of the selected task for this novel population.

Procedure

A single task was administered. The participants were asked to complete the Test for Grammatical Gender Assignment to Greek Nouns (Varlokosta, 2005), which aims to elicit the appropriate form of the nominative singular determiner for 141 items. Of these, 77 items are real nouns (e.g., *ippótis* "knight," *vrísi* "faucet," and *város* "weight") and represent all eight ICs. The remaining 64 items are novel, phonotactically conforming nouns ending in the seven different possible suffixes for Greek nouns (e.g., *péfisma*, *tagherós*, and *oviléða*). Two of the suffixes (*−os* and *−i*) are ambiguous between multiple ICs and possible gender values, while the others are unambiguous. The real and novel nouns constitute two different blocks; the order of items within each block was pseudo-randomized and fixed. The full word-list can be found in Appendix A.

The task was carried out orally, to exclude any facilitatory effect of the Greek morpho-historical spelling, and no vocabulary pre-test was carried out. Each session began with a short training phase consisting of two to four items, depending on the participant's performance. The aim was to explain to the participants that words in Greek are often accompanied by "a little friend" (the definite determiner). During the introduction, the investigator explained to each participant that "words are often not alone, but instead have a little friend that comes before them"–such as *papús–o papús* ("granddad"–"DET_M granddad"), *jajá–i jajá* ("grandma"–"DET_F grandma"), *vivlío–to vivlío* ("book"–"DET_N book"). Comprehension of the task requirements was tested by using some high-frequency pairs such as *mamá–i mamá* ("mum"–"DET_F mum") as training items before the main experiment. One participant from the younger age group who was not able to carry out the task, i.e., to provide a determiner as a response, was excluded from the study.

RESULTS

Overview

The collected data were digitized and rated by a native Greek speaker. The two groups showed clear differences in their accuracy when compared against each other and within groups when comparing ICs. **Table 2** shows the accuracy of each group per IC in the real noun condition. In **Table 3** we report the accuracy per suffix in the novel noun condition and additionally we compare their performance to that of end-state L1 speakers of Greek as reported by Varlokosta (2011) [Note that Varlokosta (2011) only administered the novel nouns sub-task as the study concerned adult monolinguals' gender assignment performance based solely on the information carried by the noun suffix]. In Appendix C, the Tables C1a–b and C2a–b present individual participant results for both groups in the respective conditions.

In order to maintain comparability with the control group's results as reported by Varlokosta, but also because only the nominative singular form of the novel nouns is presented to the participants, it is not possible to assign the novel nouns to the same ICs that we employ for the real nouns. For example, nouns ending in/os/cannot be assigned to IC1, IC7, or the exceptional imparisyllabic *−s* neuters. Therefore, the responses for all novel nouns with that phonological ending are grouped under a category "*−os* M/F/N." The feminine IC3 and IC4 and neuter IC6 categories with the phonological ending/i/are also collapsed into one category ambiguous between F and N gender values, as the null suffix that is posited for IC3 and IC4 cannot be assumed to exist in the absence of its consequences on the inflectional paradigm. The neuter suffix *−ma* is treated as unambiguous in relation to feminine nouns of the IC3 category with a null suffix and/a/as their thematic vowel (Varlokosta, 2011; but see the section "Discussion" below for a challenge). Comparison between the real and novel noun conditions is therefore less straightforward by necessity.

The performance per morphological paradigm and prototypicality class (see **Table 4**) in the real noun condition, and for each suffix in the novel noun condition, will be elaborated on in the following subsections.

TABLE 2 | Accuracy, mean number, standard deviation, and range of correct responses by child and adolescent Heritage Greek speakers (real nouns).

Inflectional Category (Items)	Children (Accuracy%, *M*, *SD*, Range)	Adolescents (Accuracy%, *M*, *SD*, Range)
IC1 *–os* M/F (12)	26.74, 3.21, 1.61, 0–6	48.08, 5.77, 2.28, 3–12
IC2 *–s* M (12)	22.57, 2.71, 2.49, 0–12	62.82, 7.54, 4.03, 2–12
IC3 *–Ø* F (15)	33.33, 5.0, 2.54, 2–12	83.08, 12.46, 2.85, 6–15
IC4 *–Ø* F (2)	14.58, 0.29, 0.55, 0–2	88.46, 1.77, 0.44, 1–2
IC5 *–o* N (11)	71.21, 7.83, 2.57, 3–11	64.34, 7.08, 3.38, 1–11
IC6 *–i* N (13)	74.68, 9.71, 2.63, 5–13	68.05, 8.85, 4.3, 0–13
IC7 *–os* N (4)	70.83, 2.83, 1.31, 0–4	42.31, 1.69, 1.18, 0–4
IC8 *–ma* N (6)	68.75, 4.13. 1.85, 0–6	51.28, 3.08, 2.53, 0–6
Imparisyllabic *–s* N (2)	81.25, 1.63, 0.58, 0–2	42.31, 0.85, 0.9, 0–2

TABLE 3 | Accuracy, mean number, standard deviation, and range of correct responses by child and adolescent Heritage Greek speakers, compared to the accuracy of typical L1 adult speakers as reported by Varlokosta (2011) (novel nouns).

Phonological Ending	Adults L1 Greek (Varlokosta, 2011)	Children Heritage Greek (Accuracy%, *M*, *SD*, Range)	Adolescents Heritage Greek (Accuracy%, *M*, *SD*, Range)
***−os (12)* as**	100	100, 12, 0, 12–12	100, 1.92, 0.28, 11–12
M	85.8 (M)	24.31 (M)	78.71 (M)
F	4.3 (F)	16.67 (F)	2.58 (F)
N	9.8 (N)	59.03 (N)	18.71 (N)
***−i* F + N (12)**	99.5 (44.0 F)	83.97, 10.04, 1.81, 6–12 (21.95 F)	92.31, 11.08, 1.12, 9–12 (60.9 F)
***−is* M (8)**	94.3	20.94, 1.67, 1.99, 0–8	51.92, 4.15, 2.67, 0–8
***−as* M (8)**	96.0	23.44, 1.88, 1.9, 0–8	59.62, 4.77, 3.17, 0–8
***−a* F (10)**	91.0	27.62, 2.75, 2.72, 0–10	75.19, 7.46, 2.76, 3–10
***−o* N (8)**	94.5	61.46, 4.92, 2.48, 1–8	50.96, 4.08, 3.09, 0–8
***−ma* N (6)**	74.3	50.69, 3.04, 1.99, 0–6	49.35, 2.92, 1.85, 0–6

TABLE 4 | Overview of the Greek prototypicality categories according to Anastasiadi-Symeonidi and Cheila-Markopoulou (2003).

	+Prototypical	−Prototypical
Masculine	+animate, male, *−s*	−animate, *−s*
Feminine	+animate, female, *−a/−i/−u*	−animate, *−a/−i*
	−animate, abstract, *−a/−i*	+animate, professional, *−s*
Neuter	−animate, *−o/−i/−a*	−animate, *−s*
	+animate, animal and human young, *−o/−i*	+animate, non-diminutive animal, *−o/−i*

The novel nouns bearing the suffix −os can be plausibly parsed either as IC1 masculine or feminine nouns or as IC7 neuter nouns, therefore any nominative determiner response is coded as accurate and only the failure to provide a determiner is coded as an incorrect response. Similarly, the novel nouns bearing the suffix −i can be plausibly parsed as IC3 or IC4 feminine nouns or as IC6 neuter nouns. A masculine determiner or the failure of providing one is coded as an incorrect response.

In the real noun condition, we observe that the older group (adolescents) performed better than the younger group (children) in ICs 1–4, which included all masculine and feminine nouns, while the children appear more accurate in the neuter ICs 5–8. In all ICs, except for two-item IC4, the child participants' performance deviates less from the group average compared to the corresponding values for the adolescent group. When examining the overall performance of individual participants, we observe that the younger group's accuracy across all tasks reached 52% (SD = 0.1), with the worst performing child achieving 30% accuracy and the best, 72%. The adolescent group's overall accuracy reached 61% (SD = 0.17), with the worst performing adolescent achieving 38% accuracy and the best, 97%.

In the novel noun condition, we observe that for the unambiguous suffixes *−is*, *−as*, *−a*, *−o*, and *−ma*, the adolescents performed better than the children in the masculine and feminine classes, while the children performed better in the neuter *−o* class. The two groups had comparable accuracy in the neuter *−ma* class, but the children performed slightly better in the neuter *−o* class. The performance of individual participants from their respective group averages presents a slightly different picture compared to that for the real nouns. The adolescent group exhibited smaller variance in the phonological endings *−i* and *−ma*. The overall child group accuracy was at 53% (SD = 0.09), with the lowest scoring participant reaching 35% and the highest scoring 73%. The overall accuracy of the adolescent group reached 68% (SD = 0.18), with the lowest scoring participant reaching 43% and the highest scoring reaching 100% accuracy.

While we are not able to compare the two heritage speaker groups to a monolingual homeland speaker control population in the real noun condition, we expect that, in line with the previous studies presented in the section "Introduction," the homeland Greek speakers will perform near perfectly or at least at the same level as their performance in the novel nouns.

Accuracy by Inflectional Class of Real Nouns

The test items from IC1 are 12 nouns ending in *−os*, equally split between masculine and feminine. The performance of the two age groups was clearly different, with child participants only being able to produce the correct determiner almost 27% of the time, while the adolescents were successful over 48% of the time. The children erroneously provided a neuter determiner for more than half of the words. The adolescents provided a masculine determiner two thirds of the time but rarely identified words as feminine, instead erroneously producing a neuter determiner one fourth of the time.

There are 12 masculine nouns in IC2 ending in *−s*, of which half are of the *−is* type and the other half of the *−as* type. Most children were unable to correctly identify these nouns as masculine, providing a neuter determiner at more than 61%, and a feminine determiner at over 15%. The adolescents were successful in 62% of cases, but provided a neuter determiner at more than 25% of the time.

The IC3 items include 15 feminine nouns ending in the feminine suffix, which is phonologically null in the nominative singular and surfaces as *−es* in the plural. Nine of the words in this category have a stem with the thematic vowel /a/, while the rest have the thematic vowel /i/. The adolescents correctly

identified these nouns as feminine in an overwhelming 83% of the time, and the majority of their erroneous responses skewed toward the neutral determiner. On the other hand, the child participants were only successful one third of the time, yet their erroneous responses also skewed toward the neuter determiner over half of the time.

From IC4, there are only two learned vocabulary items marked feminine with the null suffix in the nominative singular that manifests as *−is* in the nominative plural. Both stems end in the thematic vowel /i/ and the performance of the two groups followed the same pattern as for IC3.

IC5 is represented with 11 neuter nouns ending in *−o*. The younger group was accurate more than 71% of the time and their erroneous responses were nearly equally distributed between the other two genders (**Table 5**). The older group was accurate approximately 64% of the time, but their erroneous responses skewed overwhelmingly toward the masculine determiner.

IC6 includes 13 neuter nouns ending in *−i*. The accuracy of the two groups was comparable but slightly improved to that of IC5, with the exception of the error trend in the older group (adolescents), which this time skewed toward the feminine determiner (**Table 6**).

IC7 has four neuter nouns ending in *−os*, which follow a distinct inflectional paradigm from masculine or neuter nouns ending in *−os* from IC1. The children continued to perform as they did for IC5 and IC6, correctly providing the neuter determiner; in case of erroneous responses, their mistakes were approximately equally distributed between masculine and feminine. The adolescent participants were accurate well under half of the time and provided a masculine determiner in a majority of the cases (53%).

The six neuter nouns from IC8 are deverbal nouns ending in *−ma*. Here, too, the children mostly provided the correct neuter determiner (69%). Where they replied incorrectly, they favored a feminine (23%) over a masculine determiner (8%). The adolescents provided the correct response at a lower rate (51%). When replying incorrectly, they showed a clear preference for a feminine (46%) over a masculine determiner (3%).

Finally, in the exceptional class of the two imparisyllabic neuter *−s* nouns, the performance of our participants nevertheless does not deviate from what we observed in IC7: The children accurately assigned a neuter determiner in the vast majority of cases (81%) and their errors skewed toward the masculine determiner (15%). In contrast, the adolescents only provided a correct response in a minority of cases (42%) and showed a strong preference for assigning a masculine determiner (50%).

Accuracy by Prototypicality Class in Real Nouns

The same vocabulary items were placed in the categories proposed by Anastasiadi-Symeonidi and Cheila-Markopoulou (see **Table 4**). The nouns bearing different nominal suffixes may be prototypical or non-prototypical members of the category that the given nominal suffix defines based on semantic properties, mainly their animacy value and, on some occasions, additionally other semantic factors such as concreteness or diminution. The full word-list is provided as Appendix B. The **Tables 7–10** present the aggregate group results. In Appendix C, the Tables C3a–b present individual participant results.

TABLE 5 | Accuracy and distribution of the children's answers for the Real Nouns.

	Children			
	% accurate	% as masculine	% as feminine	% as neuter
IC1 *−os* M/F	26.74	28.1	13.1	58.6
IC2 *−s* M	22.57	22.57	15.6	61.8
IC3 *−Ø* common F	33.33	11.39	33.33	55.28
IC4 *−Ø* learned F	14.58	14.58	14.58	70.83
IC5 *−o* N	71.21	15.15	13.64	71.21
IC6 *−i* N	74.68	14.1	11.22	74.68
IC7 *−os* N	70.83	15.63	13.54	70.83
IC8 *−ma* N	68.75	8.33	22.92	68.75
Imparisyllabic *−s* N	81.25	14.58	4.17	81.25

TABLE 6 | Accuracy and distribution of the adolescents' answers for the Real Nouns.

	Adolescents				
	% accurate	% as masculine	% as feminine	% as neuter	% no answer
IC1 *−os* M/F	48.08	66.6	7.4	26	–
IC2 *−s* M	62.82	62.82	9.6	27.6	–
IC3 *−Ø* common F	83.08	2.05	83.08	14.36	0.51
IC4 *−Ø* learned F	88.46	3.58	88.46	7.69	–
IC5 *−o* N	64.34	32.17	3.5	64.34	–
IC6 *−i* N	68.05	4.73	27.22	68.05	–
IC7 *−os* N	42.31	53.85	3.85	42.31	–
IC8 *−ma* N	51.28	2.56	46.15	51.28	–
Imparisyllabic *−s* N	42.31	50	7.69	42.31	–

TABLE 7 | Accuracy, number of items, mean number, standard deviation, and range of correct responses per prototypicality condition in the younger group (children).

	Children (Accuracy%, *n*, *M*, *SD*, Range)		
	Masculine	Feminine	Neuter
+Prototypical, +Animate	45.37, 9, 4.08, 2.06, 1–9	47.92, 6, 2.88, 1.57, 0–6	70.14, 6, 4.21, 1.28, 2–6
+Prototypical, −Animate	–	22.5, 5, 1.13, 1.23, 0–4	71.3, 18, 12.83, 5–18
−Prototypical, +Animate	–	–	74.31, 6, 4.46, 1.79, 0–6
−Prototypical, −Animate	15.74, 9, 1.42, 1.89, 0–9	17.36, 12, 2.08, 2.1, 0–7	74.31, 6, 4.46, 1.72, 0–6

TABLE 8 | Accuracy, number of items, mean number, standard deviation, and range of correct responses per prototypicality condition in the older group (adolescents).

	Adolescents (Accuracy%, *n*, *M*, *SD*, Range)		
	Masculine	**Feminine**	**Neuter**
+Prototypical, +Animate	78.63, 9, 7.08, 1.8, 4–9	89.74, 6, 5.38, 0.77, 4–6	64.1, 6, 3.85, 1.63, 1–6
+Prototypical, −Animate	–	84.62, 5, 4.23, 1.09, 2–5	62.82, 18, 11.31, 4.82, 3–18
−Prototypical, +Animate	–	–	64.1, 6, 3.85, 1.95, 1–6
−Prototypical, −Animate	60.68 9, 5.46, 3.43, 0–9	44.87, 12, 5.38, 2.6, 2–12	42.31, 6, 2.54, 1.94, 0–6

TABLE 9 | Distribution of responses for novel nouns by the younger group (children).

	Children		
	% as masculine	**% as feminine**	**% as neuter**
−*os* ambiguous (M/F/N)	24.31	16.67	59.03
−*i* ambiguous (F/N)	16.03	21.95	62.02
−*is* M	20.94	16.23	62.83
−*as* M	23.44	15.63	60.94
−*a* F	17.15	27.62	55.23
−*o* N	18.23	20.31	61.46
−*ma* N	19.44	29.86	50.69

TABLE 10 | Distribution of responses for novel nouns by the older group (adolescents).

	Adolescents		
	% as masculine	**% as feminine**	**% as neuter**
−*os* ambiguous (M/F/N)	78.71	2.58	18.71
−*i* ambiguous (F/N)	7.69	60.9	31.41
−*is* M	51.62	14.42	25.96
−*as* M	59.62	14.42	25.96
−*a* F	4.65	75.19	20.16
−*o* N	44.23	4.81	50.96
−*ma* N	3.9	46.75	49.35

Prototypical animate nouns may fall into any of the three gender categories. Nine masculine, six feminine, and six neuter nouns are included in this category. Under this classification, the younger group (children) provide correct responses for approximately half of the items in the masculine and feminine conditions but had an accuracy of slightly over 70% in the neuter condition. The older group (adolescents) was very accurate in the feminine condition, approaching 90%, and also in the masculine condition with over 78% of correct responses. Their accuracy in the neuter condition was well above chance but not as remarkable as in the previous two.

The prototypical inanimate class does not include any masculine nouns but consists of five feminine and 18 neuter nouns. In this category, the younger participants performed worse in the feminine condition, with correct responses making up only slightly over one fifth of the total, while their accuracy in the neuter condition was similar to their accuracy in the prototypical animate category. The older participants' accuracy was not remarkably different compared to the previous category.

The non-prototypical animate category includes only six neuter nouns and the two groups showed comparable accuracy as in the neuter conditions in both prototypical categories.

Finally, the non-prototypical inanimate category includes nouns from all three genders, of which nine were masculine, 12 were feminine, and six were neuter. The nouns in this category elicited some of the lowest accuracy rates in the prototypicality analysis. The younger participants provided accurate responses well below one fifth of the time in the masculine and feminine conditions, but they retained their previous levels of accuracy in the neuter condition. The older participants, who previously were remarkably accurate in the feminine and masculine conditions, did not perform as accurately here, correctly responding approximately 45 and 61% of the time, in the respective categories. Their relatively lower accuracy when it comes to neuter nouns was more pronounced in this category, reaching only slightly over 42%.

When looking at the performance of individual participants, we observe higher deviations from the group accuracy rate, with extremes such as a standard deviation of 12.83 in the Prototypical Inanimate Neuter category for child participants. Overall, under the prototypicality analysis, no age group appears to consistently achieve more heterogenous results.

Accuracy by Nominal Suffix in Novel Nouns

The novel nouns, by definition devoid of semantic associations, were only analyzed in terms of morphological form. Additionally, because the participants did not know these words beforehand, homophonous suffixes that may belong to separate inflectional classes–namely *−os* (M/F vs. N) and *−i* (F vs. N)–were aggregated together, since the participants could not be aided by familiarity with inflected forms.

The *−os* category, which consisted of twelve items, was expected to elicit all three gender-marked determiners. Nevertheless, both groups showed a clear preference, albeit a different one. In nearly 60% of the cases, children preferred assigning the neuter gender to these words. The adolescents, in contrast, preferred the masculine determiner in nearly 80% of the cases. In both cases, the feminine determiner was the least popular response, although the distance between that and the other two options was more striking in the adolescents' responses.

The *−i* category included 12 items that were expected to elicit either feminine or neuter determiners. The two groups indeed preferred the two options in most cases, with masculine determiners making up only 16% of the responses in the younger group, and almost 8% in the older group. The two age

groups once again showed distinct preferences, with the children preferring the neuter option in slightly over 62% of the cases, and the adolescents the feminine in slightly under 61%.

The *−is* category consists of eight nouns that were expected to elicit only masculine determiners. The expectation was subverted by both groups, as only approximately 21% of the items elicited a masculine determiner in the younger group, and the older group provided one only slightly above 51% of the time. The distribution of erroneous responses favored once more the neuter determiner, especially in the younger group, while the older participants also provided a non-trivial amount of feminine determiners in this category.

The *−as* category has eight items, all expected to elicit masculine determiners. Both groups performed comparably to the *−is* category, with the older group correctly producing a masculine determiner in 8% more of the cases. The younger group only saw a 2.5% improvement, but still overwhelmingly preferred the erroneous neuter determiner.

The *−a* category consists of 10 items expected to elicit feminine determiners. The older participants successfully provided them in three fourths of the cases, with most of their erroneous replies favoring the neuter determiner. The younger participants provided the correct determiner in slightly under 28% of the cases, replying mostly with the neuter determiner in over 55% as well as with a non-trivial amount of masculine determiners.

The *−o* category has eight test items expected to elicit a neuter determiner. Neither group performed clearly as expected. The children produced the neuter determiner approximately 60% as it did in the previous categories, while the adolescents were almost equally split between neuter and masculine.

The *−ma* category included six items expected to elicit a neuter determiner. The two groups also defied this expectation, with the younger group accurately responding only 51% of the time, and their erroneous responses skewing toward the feminine determiner, while the older group was once more nearly equally split, this time between feminine and neuter.

DISCUSSION

By employing a task that directly elicits a gender-marked determiner, we were able to detect a difference in performance both between two age groups of heritage speakers, as well as between heritage speakers and homeland speakers where comparisons are possible.

The results of this study highlight a likely contribution of language exposure (by proxy of age) and of specifically phonological–rather than expected broadly morphological–cues in the process of gender assignment in both the real-word and the novel-word tasks. In turn, the expected facilitatory effects of semantic information as introduced by the concept of prototypical gender values were not as pronounced in the real-word task.

With regard to the contribution of morphological information, across the categorizations of both the real vocabulary items and the novel nouns, the children tended to respond with the neuter determiner. The distribution of their responses suggests that the retreat to neuter is not indicative of preserved knowledge in favor of neuter vocabulary items but rather of a retreat to a default due to uncertainty regarding the gender value that the form encodes. This finding is similar to what Alexiadou et al. (2021) observed for adolescent and adult heritage speakers of Greek in the United States on the basis of a narration task. But unlike what has been proposed in that paper, in our study, the effect cannot be attributed to difficulties with the establishment of agreement chains.

The facilitatory effect of prototypicality in this group was restricted to prototypical masculine and feminine items, but it was not strong enough to enable them to perform clearly above chance level. Animacy showed an additional facilitatory effect for the feminine items, although the absence of non-prototypical animate items in this study does not allow us to tease apart the influence of the two factors. Prototypicality effects were completely absent in the cause of neuter items, where correct responses were uniformly high. The retreat to a default supports the Tsimpli and Hulk's (2013) assessment that neuter is the learner default in Greek.

The adolescents appear to be very sensitive to phonological cues in both the real and the novel noun conditions, at the expense of expected morphological cues that are used to define the various ICs. In the real noun condition, we observe a strong tendency to interpret all /s/-ending suffixes as masculine, and secondarily as neuter, which leads to their high accuracy rate for masculine nouns; but it impedes their accuracy in the rarer feminine nouns ending in *−os*. They also appear less sensitive to the paradigm difference between M + F *−os* vs. N *−os*, which may indicate less reliable access to the inflected forms of the nouns.

At the same time, the adolescents also appear to strongly associate nominal forms ending in the vowels /i/ and /a/ with feminine gender, which manifests as remarkable accuracy in the feminine noun conditions; but it also impedes their accuracy when it comes to deverbal nouns ending in *−ma*. We hypothesize that this is due to a reanalysis of *−ma* as a stem with a thematic vowel /a/ and the feminine null nominal suffix, contrary to Varlokosta's (2011) categorization of the two as unambiguous. The sensitivity to prototypicality is strongly pronounced between prototypical and non-prototypical inanimate feminine nouns, where non-prototypical feminine nouns are correctly identified as feminine at below or slightly above chance level, while prototypical feminine nouns were consistently identified correctly, more so than any other category in this analysis. A smaller animacy effect could also be observed for non-prototypical neuter items.

More broadly, we note that neither the paradigm-focused categorization system from Ralli (1994) nor the semantic-morphological system from Anastasiadi-Symeonidi and Cheila-Markopoulou (2003) allow us to fully capture the performance patterns of the heritage speakers. With regard to Ralli's system, we have discussed an apparent reliance of the heritage speakers on the phonological endings of the test items, giving rise to ambiguities that are not expected in the production of their homeland counterparts. As this categorization system relies

on access to the full paradigm, we consider unavailability or degraded access to the full paradigm to be a possible explanation for the lower accuracy of even the older heritage speakers compared to their homeland counterparts. We believe that similar elicitation tasks that attempt to elicit different combinations of gender, case, and number values can provide the data needed to pursue this line of investigation. With regard to Anastasiadi-Symeonidi & Cheila-Markopoulou's system, we have discussed some tentative indications of increased accuracy for prototypical nouns of their respective genders. We also note that under this categorization system, the ambiguities between nouns of different inflectional classes with homophonous endings are not surprising as the system does not account for inflectional paradigm differences. We remain reserved about the significance of those observations, since the test items were not balanced across the two dimensions of prototypical and animacy.

Finally, we acknowledge again a number of limitations that arise from the preliminary nature of our study and the constraints inherent to the study of heritage populations. The small number of participants and the heterogeneity the sample exhibits did not allow us to robustly examine the observations discussed above and establish their statistical significance. However, our findings in the child group do align with other studies of learner Greek, while those in the adolescent group point to differences from the homeland group. This latter pattern is quite robust as shown in section "Results."

CONCLUSION

Our study sought, on the one hand, to establish whether a determiner elicitation task would be an appropriate tool for investigating the principles guiding grammatical gender assignment in a heritage language population. On the other hand, it aimed to provide data regarding the accuracy of this group of Heritage Greek speakers with English as their dominant language, and compare their performance to that of end-state (monolingual Greek) homeland speakers. The task does appear to be appropriate: The relatively effortless and early acquisition of the grammatical gender feature in homeland speakers is not replicated in a heritage language environment, although an increased accuracy of the adolescent group can be observed in some of the inflectional categories.

By examining the error types in the data in both the real and the novel word sub-tasks, we hoped to identify a difference in the degree of reliance upon distinct grammatical gender assignment mechanisms. In the child group, the retreat to the default neuter dominates and other gender cues have limited influence. The response patterns of the adolescent group suggest that among the possible sources of a gender value, a greater sensitivity was shown to the purely phonological properties of the inflectional suffix, giving rise to unexpected ambiguities that are not encountered in (monolingual) Greek homeland performance patterns, such as that of neuter *−ma* and feminine *−a* that would be eliminated on morphological grounds.

Finally, by utilizing Anastasiadi-Symeonidi and Cheila-Markopoulou's (2003) categorization principles based on prototypicality in our alternative analysis, we attempted to evaluate the degree in which purely semantic principles could also be at play. We could identify a strong sensitivity to a prototypicality effect only in the case of feminine vocabulary items, while a weaker sensitivity to animacy was seen in non-prototypically neuter items.

The preliminary results of this study have highlighted the relevance of access to the inflectional paradigms (case and number) of the nominal suffixes in question. This holds particularly for the reliance on strictly phonological cues by the heritage speaker population in such a way that gives rise to ambiguous gender-marking situations not found in (monolingual Greek) homeland speakers' productions. We suggest that further research in the gender assignment patterns of Heritage Greek speakers which is informed by models of organization of the mental lexicon with regard to inflectional paradigms.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Cyprus National Bioethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

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

FUNDING

We acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin. AL 554/8-1 is hereby acknowledged for supporting the data collection in Australia.

ACKNOWLEDGMENTS

We are very grateful to Monika Tsiflidis for his help with data collection. We would like to thank Gero Nagel for his help with improving the dataset presentation. We would also like to thank the reviewers for their constructive comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717449/full#supplementary-material

REFERENCES

Alexiadou, A. (2004). "Inflection class, gender and DP internal structure," in *Explorations in Nominal Inflection*, eds G. Müller, L. Gunkel, and G. Zifonoun (Berlin: De Gruyter Mouton), 21–50. doi: 10.1515/9783110197501.21

Alexiadou, A. (2017). "Gender and nominal ellipsis," in *A Schrift to Fest Kyle Johnson*, eds N. LaCara, N. K. Moulton, and A.-M. Tessier (Amherst, MA: Linguistics Open Access Publications), 11–21.

Alexiadou, A., and Müller, G. (2008). "Class features as probes," in *Inflectional Identity*, eds A. Bachrach and A. Nevins (Oxford: Oxford University Press), 101–155.

Alexiadou, A., Rizou, V., Karkaletsou, F., and Tsokanos, N. (2021). Gender mismatches in Heritage Greek. *Languages* 6:3. doi: 10.3390/languages6010003

Anagnostopoulou, E. (2017). "Gender and defaults," in *A Schrift to Fest Kyle Johnson*, eds N. LaCara, K. Moulton, and A.-M. Tessier (Amherst, MA: ScholarWorks@UMass), 23–33.

Anastasiadi-Symeonidi, A., and Cheila-Markopoulou, D. (2003). "Sinchronikes ke diachronikes tasis sto genos tis Elinikis: mia theoritiki protasi [Synchronic and diachronic trends in the Greek gender: a theoretical proposal]," in *To genos [Gender]*, eds A. Anastasiadi-Simeonidi, A. Ralli, and D. Chila-Markopoulou (Athens: Patakis), 13–56.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.

Kaltsa, M., Tsimpli, I. M., and Argyri, F. (2017). The development of gender assignment and agreement in English–Greek and German–Greek bilingual children. *Linguist. Approaches Biling.* 9, 253–288. doi: 10.1075/lab.16033.kal

Kazana, D. (2011). *Agreement in Modern Greek Coordinate Noun Phrases.* Ph.D. dissertation. Colchester: University of Essex.

Laleko, O. (2019). Resolving indeterminacy of gender agreement: comparing heritage speakers and L2 learners of Russian. *Herit. Lang. J.* 16, 151–182. doi: 10.46538/hlj.16.2.3

Mastropavlou, M. (2006). *The Role of Phonological Salience and Feature Interpretability in the Grammar of Typically Developing and Language Impaired Children*. Ph. D. dissertation. Thessaloniki: Aristotle University of Thessaloniki.

Mastropavlou, M., and Tsimpli, I. M. (2011). The role of suffixes in grammatical gender assignment in Modern Greek: a psycholinguistic study. *J. Greek Linguist.* 11, 27–55. doi: 10.1163/156658411x563685

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

Paspali, A. (2019). *Gender Agreement in Native and Heritage Greek: An Attraction Study*. Ph. D. dissertation. Berlin: Humboldt Universität zu Berlin.

Picallo, M. C. (2017). A note on the locus and function of formal gender. *Borealis Int. J. Hisp. Linguist.* 6, 1–20. doi: 10.7557/1.6.1.4097

Polinsky, M. (2008). Gender under incomplete acquisition: heritage speakers' knowledge of noun categorization. *Herit. Lang. J.* 6, 40–71. doi: 10.46538/hlj.6.1.3

Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.

Ralli, A. (1994). "Feature representations and feature-passing operations in Greek nominal inflection," in *Proceedings of the 8th Symposium on English and Greek Linguistics*, (Thessaloniki: Aristotle University of Thessaloniki), 19–46.

Ralli, A. (2002). The role of morphology in gender determination: evidence from Modern Greek. *Linguistics* 40, 519–551.

Ralli, A. (2003). "O kathorismos tu gramatiku genus sta usiastika tis neas elinikis [Determination of grammatical gender in the nouns of Modern Greek]," in *To genos [Gender]*, eds A. Anastasiadi-Simeonidi, A. Ralli, and D. Chila-Markopoulou (Athens: Patakis), 57–99.

Stephany, U. (1997). "The acquisition of Greek," in *The Crosslinguistic Study of Language Acquisition*, Vol. 4, ed. D. I. Slobin (Mahwah, NJ: Lawrence Erlbaum), 183–333.

Tsimpli, I. M., and Hulk, A. (2013). Grammatical gender and the notion of default: insights from language acquisition. *Lingua* 137, 128–144. doi: 10.1016/j.lingua.2013.09.001

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

Varlokosta, S. (2005). *Test Apodosis Gramatiku Genus Sta Usiastika tis Elinikis [Test for Grammatical Gender Assignment to Greek Nouns]*. Rhodes: University of the Aegean.

Varlokosta, S. (2011). "The role of morphology in grammatical gender assignment: a psycholinguistic study in Greek," in *Morphology and Its Interfaces*, eds A. Galani, G. Hicks, and G. Tsoulas (Amsterdam: John Benjamins), 49–76.

Varlokosta, S., and Nerantzini, M. (2013). Grammatical gender in specific language impairment: evidence from determiner–noun contexts in Greek. *Psychology* 20, 338–357. doi: 10.12681/psy_hps.23545

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fpsyg.2021.735232

English Adjectives and Estonian Nouns: Looking for Agreement?

Daria Bahtina[1†], *Helin Kask*[2†] *and Anna Verschik*[2*†‡]

[1] *Department of Linguistics, University of California, Los Angeles, Los Angeles, CA, United States,* [2] *School of Humanities, Tallinn University, Tallinn, Estonia*

This study investigated how speakers of Estonian as L1 with varying degree of proficiency in English judge grammaticality of bilingual constructions *English adjective + Estonian noun* from the point of view of adjective agreement. Estonian is rich in inflectional morphology, and adjectives agree with nouns in case and number. The empirical evidence from English-Estonian bilingual speech shows that agreement is not always the case even when an English adjective fits into Estonian declension system. It is hypothesized that the higher proficiency in/exposure to English is, the higher is the acceptability of bilingual adjective phrases, and (non-)agreement does not play a role. To test this, an experiment was designed where the test corpus of 108 sentences consisted of real and constructed examples, both in agreement and non-agreement condition. Real sentences came from fashion and beauty blogs and vlogs. The test was administered online and the participants were asked to rate adjective acceptability. The hypothesis was confirmed: increased proficiency in English, together with younger age, had a positive correlation with acceptability of all adjective types, independent of adjective (non-)agreement. Residence and birthplace had a small effect on acceptability of some adjective types. Whether sentences were real or constructed, had only a minor effect. Male participants tended to assess real sentences lower, probably because of the topics typical for female blogs. Monosyllabic consonant-ending adjectives were exceptional, as their assessment did not depend on any factor. All in all, the study demonstrated that grammaticality judgment among the native speakers of the same L1 differs because of different degrees of bilingualism, and structural factors, such as compatibility with Estonian declension system, are not decisive. Thus, it is not clear what an ideal native speaker is.

Keywords: language contacts, bilingual constructions, grammaticality judgement, Estonian, English, multilingualism, morphosyntactic integration

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez, University of the Balearic Islands, Spain

Reviewed by:
Valérie Saugera, University of Connecticut, United States
Takeshi Nozawa, Ritsumeikan University, Japan
Sunil Kumar Ravi, Shravana Institute of Speech and Hearing, India

***Correspondence:**
Anna Verschik
annave@tlu.ee

†*These authors share first authorship*

‡*Senior author*

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

Received: *02 July 2021*
Accepted: *14 September 2021*
Published: *14 October 2021*

Citation:
Bahtina D, Kask H and Verschik A (2021) English Adjectives and Estonian Nouns: Looking for Agreement? Front. Psychol. 12:735232. doi: 10.3389/fpsyg.2021.735232

INTRODUCTION

The interest toward how a non-native speaker differs from a native speaker and whether native(-like) competence can be achieved is explicitly stated in the context of SLA/bilingualism/multilingualism research with the focus on languages that are acquired later than L1 (i.e., L2 etc.). Since the 1980s the notion of a monolingual native speaker as an ideal and a yard stick has been questioned by some scholars (e.g., Skutnabb-Kangas, 1984; Rampton, 1990) because it is not always clear what "native" is: at times, the language acquired first is not one's dominant language, and internal identification may not coincide with external (that is, how other

speakers assess a given speaker). Piller (2002) showed that so-called non-native speakers may pass for native speakers, as quite a lot depends on whether those who assess have a prior knowledge of the fact that the language in question was not L1 for the person whom they assess. In the same spirit, Gnevsheva (2017) noted that sometimes so-called non-native speakers may be perceived as speakers of different regional varieties.

Cook (2002) and Dewaele (2017) suggest the term "language user," i.e., L2, L3 or, as Dewaele (2017) has it, LX user rather than learner because languages may be acquired in a variety of ways, also without explicit learning. The notion of multicompetence, introduced in the early 1990s (see more in Cook, 2016) was a useful contribution to the field because it shed light on the monolingual bias. First, a multilingual user is not a sum of several monolinguals, and a multilingual speaker is not an imperfect version of a monolingual speaker (Murahata et al., 2016; Dewaele, 2017). Second, multilinguals have several linguistic systems in their mind, which renders their cognition and perception different from those of monolinguals who have only one linguistic system. Therefore, the latter are unable to juxtapose and draw parallels between the systems. It implies, among other things, that focusing on a particular LX without considering all other languages of a given individual gives a rather patchy picture. Important as this may be, the debate on native competence is centered around languages acquired later than L1, such as SLA, language pedagogy, teachers who are or are not native speakers of the language they teach and so on.

Quite remarkably, contact linguistics was not a part of the debate, probably because its scope and purpose are different from that of SLA and, at times, of bilingualism/multilingualism research. First, contact linguistics is not concerned with ways to achieve target acquisition; by definition, the discipline is concerned with contact-induced language change. If non-target acquisition occurs, in contact linguistics this is not relevant for comparison with an ideal native speaker; the focus is rather on the fate of this new variety (becoming an ethnolect/in-group register, diffusion into the mainstream, and so on). Multilingual communities have their own norms and what looks like imperfect acquisition from a synchronic point of view may have significant effects on the mainstream variety in a diachronic/historical perspective ("incomplete" acquisition of Baltic by speakers of Finnic that yielded Latvian is a textbook example, see Thomason and Kaufman (1988, p. 239) and references therein). In other words, the mistakes of today are the grammar of tomorrow.

Second, contact linguistic research often puts centerstage changes in L1, while SLA deals with L2. Thus, the other side of the coin is the change of perception of L1 among so-called native speakers with some proficiency in L2 (LX). For instance, research on Netherlands Turkish (Dogruöz and Backus, 2009) has demonstrated that some contact-induced features that have emerged under the impact of Dutch are a new norm, while for speakers of Turkish in Turkey these features appear erroneous and, therefore, non-native.

This discrepancy between the contact linguistics view on language change, including L1, and the notion of a monolingual native speaker, still common in SLA research, is the starting point for this article. The focus is on perception of adjective-noun (non-)agreement in bilingual phrases English adjective + Estonian noun, for instance *fancy-d kinga-d* (fancy-PL.NOM shoe-PL.NOM) "fancy shoes." Estonian is agglutinating-fusional language with highly developed inflectional morphology, where adjectives agree with nouns in case and number. From a structural point of view, English adjectives that fit into Estonian declension system would take on Estonian inflections. However, it is not always the case (Kask, 2019). We assume factors other than structural compatibility play a role here.

A number of approaches to multilingual language use focus on formal distinctions and classify phenomena accordingly (i.e., lack of integration = code-switching and integration = borrowing). Albeit we do not agree with the formal constraints proposed in the Matrix Language Frame model (Myers-Scotton, 1997), we believe that the description of the following options is empirically right: a stem from Embedded Language may retain inflections from that language, it can take on inflections from the Matrix Language (i.e., the main language of the clause), and it can remain without any markers (so-called bare forms). A code switched item can become a borrowing, if it is useful and if it gains currency and becomes a new norm. Neither is this project concerned with conventionalization (and codification) in monolingual use, as customary in Anglicisms research. It will be discussed below that multilingual speakers do not necessarily need the same mechanisms of integration as monolinguals do. In terms of classification, we can use Muysken's (2000) typology, in which English adjectives can be treated as insertional code-switching, as opposed to alternational code switching and congruent lexicalization. We concur with usage-based approaches to contact-induced language change (Zenner et al., 2019) in general and to borrowing in particular (Backus, 2012, 2014). In these approaches, the difference between one-word code-switching and borrowing cannot be described in formal terms. Instead, it is rather a matter of frequency and conventionalization than any formal criteria. This is why our underlying research question is centered around the acceptability of various English adjectives as a harbinger of new norms in the making.

Estonia is a small country with a population of 1.3 million, of which speakers of Estonian as L1 make up roughly 68%. Multilingualism is not new in Estonia, German and Russian have been sociolinguistically dominating languages for centuries (Ariste, 1981). Language contacts between Estonian and English are relatively recent, starting from the 1990s after the restoration of independence, and the changes are on-going. Recent decades have witnessed growing competence in English among younger Estonian-speakers: 84% of Estonians in the age group 15–24 and 64% in the age group of 25–39 claimed their ability to speak English (Kruusvall, 2015). According to the Housing and Population Census 2011, 38.3% of Estonian residents have indicated their knowledge of English, and 20% claim they do not speak English. However, research has shown that the percentage of Estonians who are not proficient in English is decreasing and the percentage of Estonians who have active knowledge of English is rising (Koreinik and Tender, 2013, p. 86; Kruusvall, 2015, p. 78). Students consider English as the most important subject at school (Tammemägi and Ehala, 2012, p. 249), and

96.5% of Estonian students in upper secondary general education learnt English as a foreign language (Eurostat, 2017). In addition, English is quite often also acquired informally *via* the internet and popular culture. The differences in linguistic resources between generations in Estonia is quite obvious. Thus, it may be assumed that Estonians who do not use English and those who use English on a regular basis and produce multilingual speech might have different kinds of linguistic awareness.

We employ the usage-based approach to contact-induced language change, in particular, cognitive contact linguistics (Backus, 2014; Zenner et al., 2019), assuming that there is no dichotomy between competence and performance, and competence is shaped by usage. In such approaches, language use and cognition are interconnected: new circumstances affect language use that in its turn affects cognition and perception of norms. In other words, grammar is neither pre-programmed nor static but is shaped by usage. This is in line with the above-mentioned idea that a multigual's knowledge of their L1 differs from that of a monolingual because input, output, linguistic resources, communicative patterns etc. are different. Taking this into consideration, we assume that those who use English more would be less affected by (in)compatibility of English adjectives with the Estonian declension system and would accept bilingual adjective phrases regardless of the shape of adjectives.

In order to check our assumption, we designed an experiment where we tested perception of real and constructed sentences containing English Adj + Estonian N, both in the agreement and the non-agreement conditions, among various speakers of Estonian as L1 (see Data and Method section). One of the reasons why we opted for an experimental study is an almost complete lack of Estonian-English bilingual speech corpora (our own corpus of bilingual blogs and vlogs is respectively 275,000 and 48,000 words)[1]. By bilingual blogs and vlogs we mean that all entries contain either overt elements from English (insertional or alternational code-switching) or loan translations, constructions, patterns etc. More on Estonian-English blogs and vlogs see in Kask (2019, 2021), Verschik and Kask (2019). Personal blogs and vlogs are not restricted with the notion of "correctness" and, therefore, give a picture of naturalistic language use.

Our research questions are as follows:

(1) Whether there is a difference in grammaticality judgement depending on proficiency in and exposure to English;
(2) whether there is a difference in perception of real and constructed examples;
(3) whether there is a difference in perception of the group of English adjectives that fit into Estonian declension system but do not have Estonian inflections in our corpus.

The article is organized as follows. In Adjectives in bilingual phrase *English Adj + Estonian N* section, we describe adjective declension types in Estonian and the findings of the only existing empirical study by Kask (2019). Data and Method section focuses on the experiment methodology. The results are presented in Results section, followed with discussion and conclusions in Discussion section.

ADJECTIVES IN BILINGUAL PHRASE *ENGLISH ADJ + ESTONIAN N*

English and Estonian are typologically different: English is an isolating analytic language, while Estonian is an agglutinative language with fusional tendencies (Erelt, 2007, p. 7). Estonian has a rich inflectional morphology with 14 grammatical cases. In Estonian adjectives agree with nouns both in number and case, for instance: s*uur-te-sse maja-de-sse* "into big houses" (big-PL-ILL house-PL-ILL) where both the adjective and the noun are in the plural illative. In four cases such as terminative, essive, abessive and comitative, the agreement is in number only, the adjective is present as a stem (technically, the genitive stem, from which all oblique cases, except the partitive, are formed): *suure maja-ga* big.GEN house-COM "with a/the big house," *suur-te maja-de-ga* big-PL.GEN house-PL-COM "with (the) big house." However, English lacks grammatical cases altogether, and adjectives do not agree with nouns in number: *into big house-s* (big.SG house-PL) the noun is in the plural but the adjective remains in the singular.

Based on that, it would be expected that in a bilingual phrase English adjectives would agree with Estonian nouns, at least in the event when an English adjective is compatible with Estonian adjective declensions. A study by Kask (2019) showed that in bilingual blogs and vlogs the English adjective does not agree with the noun if it does not fit into the Estonian declension system. There is a tendency for agreement, if English adjectives are phonotactically similar to Estonian adjectives and, therefore, are structurally compatible with the declension system. Yet there are instances where the English adjective is compatible with the Estonian declension system but, contrary to expectations, does not receive Estonian inflections.

In the empirical data discussed in Kask (2019), seven types of English adjectives emerged, the types are presented in **Table 1**. Out of these, five types fit into Estonian declension system, so it would be expected that these adjectives agree with the Estonian noun. However, types marked with asterisk showed variation and there were several instances of the adjective being not integrated. Consider Examples (1) and (2) with the adjective *basic* used by two vloggers (Kask, 2019, p. 93), where *basic* can be declined as *lapik* "flat".

(1) Agreement

3 basicu-t asja, mida su-l vaja
3 basic-PART thing.PART that you.SG-ADES need
"3 basic things you need"

[1]There are various corpora of Estonian, for instance, of periodicals and fiction and of colloquial Estonian (closed to general public). The former contain edited text (from the point of view of a monolingual norm), the latter comprise phone conversations. In either of them, synchronic contact-induced language change phenomena are not annotated. The compilers of the corpora and researchers working with them are not interested in contact linguistics and multilingualism research. Estonian linguistics research community is tiny and there are less than a dozen researchers who work in the field of contact linguistics, so our own corpus is the only one of a kind.

TABLE 1 | Types of English adjectives presented with Estonian prototypes and sentence examples.

Type	English example	Estonian prototype	Sentence example
(A) Types that fit into the Estonian declension system and agree with the noun			
Monosyllabic, consonant ending	*deep*	*hell* "tender"	*Olin juba oma peas järgmist* ***diipi blogipostitus-t*** (deep.PART blogpost-PART) *kirjutamas.* "I was already writing the next deep blogpost in my head"
Disyllabic, ending with [i]	*fancy*	*tubli* "diligent"	*Imetlesime* ***fancy-sid tänava-i-d*** (fancy-PART.PL street-PART-PL). "We were admiring fancy streets."
Disyllabic in nominative, ending with [k]*	*basic*	*lapik* "flat"	*Need on üsna* ***basicu-d teksa-d*** (basic-NOM.PL jeans-NOM.PL) "These are pretty basic jeans"
Monosyllabic, ending with obstruent, subject to stem alternation*	*flat*	*pikk* "tall"	***Flati-d rehvi-d*** (flat-NOM.PL tire-NOM.PL) *mind eriti kaugele ei sõiduta.* "Flat tires won't take me very far."
Minimally disyllabic, ending with consonant both in writing and in pronunciation*	*random*	*ilus* "pretty"	*Mul on väga* ***randomi-d ehte-d*** (random-NOM.PL jewlery-NOM.PL) *kõrvades hetkel.* "I'm wearing very random jewelry in my ears right now."
(B) Types that fit into Estonian declension system but do not agree with the noun			
Ending with vowel in spelling but with consonant in pronunciation	*beige*	*kõrb* "dun"	*Loosin välja ühe* ***beige nokatsi*** (beige cap.GEN). "I will give away a beige cap."
Ending with [v] in pronunciation and are therefore similar to Estonian present participles	*impressive*	*hariv* "educational"	*Siinkohal tahaks teha sellise* ***appreciative momendi*** (appreciative moment.GEN). "Here I would like to have an appreciative moment"

Panel (A) shows adjectives that agree with nouns and panel (B) shows adjectives that do not; adjective types marked with an asterisk (*) show variation.

(2) Non-agreement

jätka-te oma basic eluviisi-ga
continue-2PL own basic lifetyle-COM
"continue with your basic lifestyle"

cf. theoretically possible *basicu eluviisi-ga*
basic.GEN lifestyle-COM
"with (your) basic lifestyle"

In Example (1), the adjective *basic* receives Estonian partitive marker, while in Example (2) it remains unmarked.

Test sentences in this experiment were constructed based on the types described in **Table 1**, which presents types of English adjectives as far as their compatibility with the Estonian declension system is concerned. Panel (A) showcases adjective types that fit and agree with Estonian nouns, panel (B) shows types that do not. Some types are in principle compatible, yet according to the empirical evidence they either exhibit variation or do not agree with nouns (Kask, 2019, p. 102, 106–115), these adjective types are marked with an asterisk (*).

DATA AND METHOD

Everyone who considered themselves an L1 speaker of Estonian qualified as a respondent. It turned out that all respondents had at least some proficiency in English. The condition was mentioned in the introductory part of the questionnaire, and 568 persons responded. In the introductory part it was emphasized that there is no right or wrong answer, and the respondents should just choose the answer that they consider appropriate. Participants who skipped the first six background questions were discarded from the analysis, which led to 401 respondents considered in this paper. An overview of the sociodemographic information can be found in **Supplementary Appendix A** (general composition) and **Supplementary Appendix B** (English proficiency); other relevant info is presented in Sociodemographic Basics section.

The questionnaire was created *via* SurveyMonkey and was accessible online in the period from February 19 to March 8, 2021. The link was distributed *via* Tallinn University School of Humanities Facebook page as well as by the authors through their social networks. The background information for each participant was collected *via* questions about age, gender, and place of birth and residence; the latter were categorized into major cities, Tallinn and Tartu, smaller urban areas, rural areas, and abroad. The survey contained a number of questions targeting self-reported proficiency in English (comprehension and production), active language use (speaking and writing), and more passive exposure (listening and reading). We provide an English translation of the survey questions and answer options in **Supplemenatry Appendix C**. Individual scores on these parameters had a significant correlation and were therefore normalized to be used jointly. We refer to this multicompetence factor as *proficiency in and exposure to English* or *English* for brevity.

The set of sentences given for assessment was composed in the following way. For every type described in **Table 1** we selected

a real sentence from the corpus described in Kask (2019) and constructed a counterpart: if the adjective in the real sentence agreed with the noun, we presented the same sentence with a non-agreeing adjective, and vice versa. Since our aim was to have multiple test sentences for every adjective type, we constructed additional sentences in both agreement and non-agreement versions (5–10 sentences per type), arriving at a total of 108 sentences presented to the respondents. There were no time constraints for experiment completion: participants used a self-paced method to respond to each stimulus. All sentences were randomized during the experiment to eliminate any possible methodological biases.

Table 2 demonstrates how the set of test sentences were composed with a reference to the example of monosyllabic consonant-ending adjectives (Estonian prototype *hell* "tender"). Real examples (R) are preceded by an asterisk (*). All other sentences are constructed (C). For the sake of comparison, another set of test sentences is provided in **Supplementary Appendix D**.

The reason for testing both real and constructed sentences is the need to have some point of comparison. If all sentences were constructed, our results would speak only of metalinguistic awareness and would not cover instances where certain utterances were attested in real usage but mostly rejected by the respondents. This method was used by Verschik (2006) where references to other experimental studies in contact linguistics can be found.

The sentences were coded according to categories they represent: adjectives were divided into agreement vs. non-agreement categories and sentences were divided into real vs. constructed. Each sentence code also contained information about the more specific adjective type. For example, *MS_C_A_R* means that the adjective in the sentence belongs to the type "monosyllabic, consonant-ending," the adjective agrees with the noun and the sentence is real. Acceptability judgements were collected using the Likert scales: respondents were asked to rate sentences on a scale from 0 ("nobody talks like this") to 4 ("I would say this myself"). We present results as an aggregate based on each sentence category. The descriptive and inferential analyses were conducted using the open source statistical environment R (R Core Team, 2020), RStudio Version 1.4.1106, and the package lme4 (Bates et al., 2012).

TABLE 2 | Test sentence examples for monosyllabic consonant-ending adjectives in the agreement (A) and the non-agreement (N) conditions.

Type	Agreement (A)	Non-agreement (N)
Monosyllabic adjectives ending in consonant (MS_C), Estonian prototype *hell* "tender"	* *Lindexis olid coolid päiksekad müügil.*	*Lindexis olid cool päiksekad müügil.*
	cooli-d päikseka-d	cool päikseka-d
	cool-PL.NOM sunglass-PL.NOM	cool sunglass-PL.NOM
	"there were cool sunglasses on sale at Lindex"	
	Selle pintsliga saab ilusa cleani tulemuse.	*Selle pintsliga saab ilusa clean tulemuse.*
	ilusa cleani tulemuse	
	nice.GEN clean.GEN result.GEN	ilusa clean tulemuse
	"with this brush you get a nice clean result"	nice.GEN clean result.GEN
	Oma chilli olekuga jäi ta kohe kõigile meelde chilli olekuga.	*Oma chill olekuga jäi ta kohe kõigile meelde chilli olekuga.*
	chilli oleku-ga	chill oleku-ga
	chill.GEN appearance-COM	chill appearance-COM
	"with his/her chill appearance s/he was remembered by everyone"	
	Müün netioksjonil oma coole riideid.	*Müün netioksjonil oma cool riideid.*
	cool-e riide-i-d	cool riide-i-d
	cool-PART.PL clothes-PL-PART	cool clothes-PL-PART
	"I am selling my cool clothes at a web auction"	
	Filmin täna vlogi ühel väga funil teemal.	*Filmin täna vlogi ühel väga funil teemal.*
	funi-l teema-l	fun teema-l
	fun-ADES topic-ADES	fun topic-ADES
	"Today I am shooting a vlog on a very fun topic"	
	Freshid joogid on pärast trenni nagu rusikas silmaauku.	*Fresh joogid on pärast trenni nagu rusikas silmaauku.*
	freshi-d joogi-d	fresh joogi-d
	fresh-NOM.PL drink-NOM.PL	fresh drink-NOM.PL
	"fresh drinks after a workout is just what the doctor ordered"	
	Sel hooajal on moes warmides toonides kudumid.	*Sel hooajal on moes warm toonides kudumid*
	warmi-de-s tooni-de-s	warm tooni-de-s
	warm-PL-INES tone-PL-INES	warm tone-PL-INES
	"this season knitwear in warm tones is in vogue"	

Real examples (R) are preceded by an asterisk (*).

RESULTS

Sociodemographic Basics

Given the fact that the participants in our study were recruited without prior screening, we ran a series of preparatory tests to make sure that certain correlations between participant characteristics do not lead to erroneous interpretations.

First, we looked at the correlation between exposure to and experience in English vs. age. These two factors overlapped to a high extent: the younger the participant, the more English they seemed to have. Yet, when A Kendall's tau-b correlation was run to compare the two factors—age and English—only a moderate correlation was established ($\tau_b = 0.26$, $p < 0.01$), which verified that the two factors can be used interchangeably only in about a quarter of the data, which means that age and English still affected the outcome separately. In other words, even if the most common participant profile is that of a young and fluent person, these two factors should not be collated for the purposes of data analysis.

We then investigated the relation between the experience of living abroad and English: the longer a participant spent abroad, the more likely they were to have a higher score in overall English, $F_{(1, 396)} = 39.84$, $p = 0$. Yet, living abroad cannot be treated as a precondition for an advanced English score in our dataset as there were participants with the highest score in English and no experience of living abroad.

There were several significant associations between acceptability rates and some of the sociodemographic characteristics that were true only for a subset of test sentences: residency, gender, and marginal associations for birthplace. We present the results of these factors in this section; other factors—age and English—showed more consistency within categories and adjective types and will be presented in Real vs. Constructed Sentences, Category and Acceptability Rates, Adjective Type and Acceptability Rates, Multivariate Analysis with Mixed Effects Modeling and What Can(not) be Explained by Sociodemographics and Structural Compatibility sections.

Residence played a role in acceptability rates of real sentences and some adjective types. A series of *post-hoc* Tukey's HSD tests revealed that in most cases it was Tallinn residents who had significantly higher acceptability rates than residents in the rural areas. There was a significant difference by residence in the real sentences, $F_{(4, 394)} = 2.89$, $p = 0.02$. According to the *post-hoc* test, Tallinn residents provided significantly higher acceptability rates for real sentences than the rural residents, $p = 0.02$, 95% C.I. = 0.04, 0.71. A significant difference between residency categories was also observed in the non-agreeing disyllabic adjectives ending with a consonant (*basic*), $F_{(4, 394)} = 2.89$, $p < 0.01$, and the non-agreeing disyllabic adjectives ending with an [i] (*fancy*), $F_{(4, 394)} = 3.44$, $p < 0.01$. The *post-hoc* test showed that the residents of Tallinn had a higher acceptability of non-agreeing disyllabic adjectives ending with a consonant (*basic*) when compared to the residents of rural areas, $p < 0.01$, 95% C.I. = 0.11, 0.72, and those currently residing abroad, $p = 0.03$, 95% C.I. = 0.02, 0.83. For the disyllabic adjectives ending with [i] (*fancy*), Tallinn residents had significantly higher acceptability rates than rural residents, $p < 0.01$, 95% C.I. = 0.09, 0.74. Finally, there was a significant difference by residency in the agreeing monosyllabic adjectives ending with consonant (*deep*), $F_{(4, 394)} = 3.30$, $p = 0.03$. The pattern was the same, Tallinn residents had significantly higher rates than rural residents, $p < 0.01$, 95% C.I. = 0.08, 0.63.

The birthplace of respondents also had a relationship with acceptability rates of some adjective types. The disyllabic adjectives ending with [i] (*fancy*) in the agreement condition demonstrated a difference across birthplace categories, $F_{(4, 394)} = 2.70$, $p = 0.03$. The other adjective type that had a significant difference by birthplace was the agreeing monosyllabic type ending with consonant (*deep*), $F_{(4, 394)} = 3.76$, $p < 0.01$. Respondents born in Tallinn rated this adjective type higher than respondents born in smaller Estonian towns, $p = 0.02$, 95% C.I. = 0.03, 0.61. Respondents born in Tartu also rated this adjective type higher than respondents born in smaller Estonian towns, $p = 0.03$, 95% C.I. = 0.03, 0.70.

Gender had a significant relationship with the acceptability rates of real sentences (as collected from the actual blogs and vlogs) but no other sentence category (i.e., constructed or agreement and non-agreement). Real sentences received lower acceptability rates from participants who identified as males (M = 1.55, SD = 0.86) than females (M = 1.83, SD = 0.82), $t_{(395)} = 2.36$, $p = 0.02$, $d = 0.33$; this effect was there regardless of age.

Gender also affected acceptability of the monosyllabic obstruent ending adjectives: female-identifying participants had higher acceptability rates than male-identifying participants for this adjective type both in the agreement and non-agreement conditions, $t_{(395)} = 2.22$, $p = 0.03$, $d = 0.30$ and $t_{(395)} = 3.343$, $p < 0.01$, $d = 0.48$. In the agreement condition, their mean scores were M = 1.88, SD = 0.83 for females and M = 1.61, SD = 0.91 for males; in the non-agreement condition, the means were M = 1.97, SD = 0.88 and M = 1.55, SD = 0.76, accordingly.

Real vs. Constructed Sentences

The real sentences received somewhat higher acceptability ratings (M = 1.78, SD = 0.83) than the constructed sentences (M = 1.47, SD = 0.69). A Welch two-samples *t*-test showed that the difference was statistically significant but the effect size was rather moderate, $t_{(774.21)} = 5.74$, $p < 0.01$, $d = 0.40$. This difference is visualized in **Figure 1**.

Category and Acceptability Rates

When analyzing the data by category (i.e., all adjectives, real and constructed sentences as well as sentences containing agreeing and non-agreeing adjectives), there was a strong correlation regardless of the category: the combined measure for English was positively correlated with acceptability rates and participants' age was negatively correlated with acceptability rates. These results are presented in **Table 3**.

In other words, the higher the proficiency in and exposure to English, the more likely a participant to give a higher rate to adjectives in this data set. This was true for all categories of adjectives, whether the sentence was real or constructed or if the adjective agreed with the noun or not. The relationship was the opposite for age, meaning that older speakers systematically had a lower acceptability

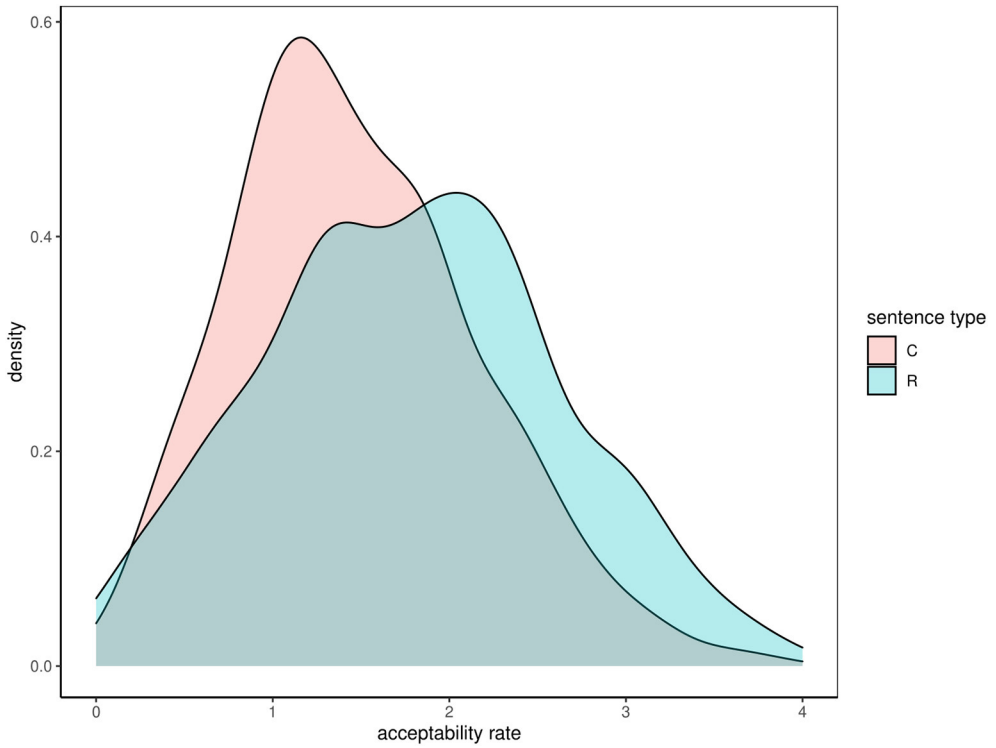

FIGURE 1 | Difference between acceptability rates for real (R) and constructed (C) sentences.

TABLE 3 | Regression results between acceptability rates across categories and two factors: proficiency in/exposure to English and age.

Category	All adjectives	Real sentences	Constructed sentences	Agreeing adjectives	Non-agreeing adjectives
English	0.23***	0.38***	0.21**	0.24***	0.22***
	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
	No. of observations			400	
Age	−0.22***	−0.32***	−0.20***	−0.23**	−0.20***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	No. of observations			399	

*The table shows coefficients and significance codes (0 "***" 0.001 "**" 0.01 "*", 0.1 "). Standard deviations are reported in parentheses.*

rate whereas younger speakers demonstrated a reliably higher acceptability rate.

Adjective Type and Acceptability Rates

We also examined acceptability rates for separate adjective types and the correlation was largely the same: most adjective types received higher rates from respondents that were younger and more proficient in and exposed to English.

As shown in **Table 4**, all adjective types except one followed that pattern. The only exception was found for adjectives ending in a monosyllabic consonant (e.g., *deep*) when it was not in agreement with the noun. Neither age nor English as factors were able to account for varying acceptability rates of this specific adjective type. We discuss this adjective type in more detail in What Can(not) be Explained by Sociodemographics and Structural Compatibility section.

Multivariate Analysis With Mixed Effects Modeling

Bivariate analyses show the association of acceptability rates with, on the one hand, respondents' age and, on the other hand, respondents' proficiency in and exposure to English. Our multivariate analysis uses lme4 package (Bates et al., 2012). This statistic assesses the ability of multiple independent variables (age and English) to predict—positively or negatively—a scalar dependent variable (acceptability rates of various categories of adjectives), while controlling for possible confounding effects, namely current residence and gender. Given the relatively small corpus sample (401 respondent), the model would not converge with more random effects, and following our usage-based approach, we selected current residence over birthplace. Our goal was to verify whether the main factors identified with the help of bivariate

TABLE 4 | Regression results between acceptability rates across adjective types and two factors: proficiency in/exposure to English and age.

	DS_C_A	DS_C_N	DS_i_A	DS_i_N	DS_k_A	DS_k_N	MS_Obs_A
English	0.22***	0.39***	0.38***	0.16*	0.27***	0.30***	0.20**
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
	No. of observations				400		
Age	−0.21***	−0.32***	−0.31***	−0.13***	−0.28***	−0.29***	−0.19***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
	No. of observations				399		
	MS_Obs_N	**MS_C_A**	**MS_C_N**	**V_C_A**	**V_C_N**	**V_A**	**V_N**
English	0.26***	0.22***	0.00	0.18**	0.15*	0.20***	0.28***
	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.05)	(0.06)
	No. of observations				400		
Age	−0.23***	−0.19***	−0.02	−0.20***	−0.17***	−0.22***	−0.25***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	No. of observations				399		

*The table shows coefficients and significance codes (0 "***", 0.001 "**", 0.01 "*", 0.1 " "). Standard deviations are reported in parentheses.*

TABLE 5 | Linear mixed-effects model between acceptability rates across adjective types with two fixed effects (age and English in interaction) and two random effects (residence and gender).

Category	All adjectives	Real sentences	Constructed sentences	Agreeing adjectives	Non-agreeing adjectives
English∼age	English*	English*	English*	English*	English*
	EST = 0.402, *SE* = 0.156	EST = 0.373, *SE* = 0.179	EST = 0.365, *SE* = 0.156	EST = 0.407, *SE* = 0.158	EST = 0.396, *SE* = 0.163
	English∼age*		English∼age	English∼age*	English∼age*
	EST = −0.084, *SE* = 0.038		EST = −0.084, *SE* = 0.038	EST = −0.084, *SE* = 0.039	EST = −0.082, *SE* = 0.040
Residence	VAR = 0.005, *SD* = 0.070	VAR = 0.007, *SD* = 0.085	VAR = 0.004, *SD* = 0.067	VAR = 0.004, *SD* = 0.068	VAR = 0.005, *SD* = 0.072
Gender	VAR = 0.005, *SD* = 0.070	VAR = 0.055, *SD* = 0.233	VAR = 0.004, *SD* = 0.065	VAR = 0.006, *SD* = 0.078	VAR = 0.007, *SD* = 0.085

*The table shows significance codes for each statistically significant predictor (0 "***", 0.001 "**", 0.01 "*", 0.1 " "), estimates (EST) and standard errors (SE). We also report variance (VAR) and standard deviation (SD) for random effects.*

tests are still valid when random effects are included in the analysis.

Table 5 summarizes the main results of this linear mixed-effect model, testing the rate of acceptability outcome as predicted by age and English. The results show that both age and English are statistically significant predictors of the outcome, with opposite effects. Belonging to an older age band slightly decreases acceptability rates while English is positively associated with acceptability outcome.

We did not expect the model to work for smaller subcategories due to lower counts but some specific sentence types showed a similar pattern: increased score in English was a predictor of higher acceptability rates, belonging to an older age band predicted the opposite. More specifically, age was a significant predictor for DS_C_N, English for DS_i_A and DS_k_N, and all three (English, age and their intercept) strongly predicted the outcomes for V_A and V_N adjective types. The following types did not show any significant relation with age and English when residence and gender were taken into account: V_C_N, MS_Obs_A, MS_Obs_N, as well as MS_C_A that did not reveal any patterns in the bivariate tests either. Other adjective types did not have sufficient variance to be included in the model (i.e., singularity warning). It should be noted that while these results were statistically insignificant or omitted due to singularity warning, the directionality was always the same.

Overall, multivariate analysis corroborates findings from bivariate tests in the following way: other factors being equal (i.e., residence and gender), age has a slightly negative correlation with acceptability rates of all adjectives while English is positively correlated. These results are visually represented in **Figure 2**: acceptability rates increase with English and slightly decrease with age even when English is high.

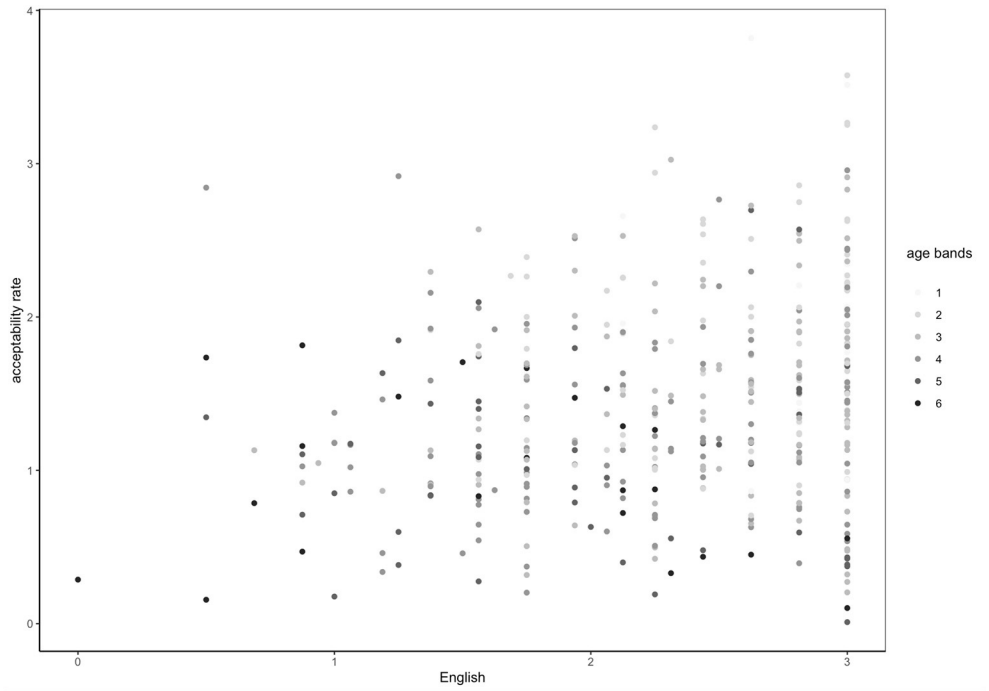

FIGURE 2 | Acceptability rate of all adjectives by exposure to and proficiency in English by age bands (1 = under 20, 2 = 21–30, 3 = 31–40, 4 = 41–50, 5 = 51–60, 6 = 61–...).

What Can(not) Be Explained by Sociodemographics and Structural Compatibility

Our data yielded a very consistent and statistically significant acceptability pattern: all test-sentences were more likely to be accepted by participants who were younger and who had a higher proficiency in and exposure to English. **Figure 3** demonstrates this pattern with a reference to one specific adjective type: disyllabic adjectives ending with [i] (*fancy*) in the agreement condition. In the study by Kask (2019) this was the type that always agreed with the noun in number and case, however, among our participants its perception exhibited variation. As "Category and Acceptability Rates" and "Adjective Type and Acceptability Rates" sections demonstrate, the acceptability rate is higher among participants with higher proficiency in and exposure to English. For example, higher levels of English significantly predicted higher acceptability rates for non-agreeing disyllabic adjectives ending with a consonant (*basic*), $b = 0.39$, $t_{(398)} = 6.08$, $p < 0.001$, R-squared = 0.09 (**Figure 3**). Younger age significantly predicted higher acceptability rates for that type of adjective, $b = -0.32$, $t_{(399)} = 10.42$, $p < 0.001$, R-squared = 0.21 (**Figure 4**). Across the data, this standard pattern was true for all types in both the agreement and the non-agreement conditions, except for one adjective type, to be discussed below in this section.

On the other hand, there was one adjective type that did not follow that or any other pattern. Monosyllabic adjectives ending with a consonant (*deep*) did not significantly interact with any factor we tested for in either the agreement or the non-agreement condition. **Figures 5** and **6** illustrate the results in the non-agreement condition. Such results contradict expectations also because the adjective *deep* is a conventionalized borrowing in Estonian (spelled *diip*) and is included in the prescriptive Estonian dictionary (Eesti Õigekeelsussõnaraamat ÕS, 2018), which suggests that the adjective would take on Estonian inflectional morphology in the same manner as non-borrowed adjectives.

DISCUSSION

We looked at the perception of (non-)agreement in bilingual adjective phrases and expected that perception of agreement/non-agreement and real/constructed examples differs among "native speakers" of Estonian. We assumed that acceptability of both agreeing and not agreeing adjectives will be higher in respondents with a high proficiency in/exposure to English. Generally speaking, the hypothesis was confirmed; still, there are some additional factors and some odd cases.

Sociodemographic Basics

Several sociodemographic observations and factors proved to have a significant impact on the answers.

The first observation is the recurrent combination of young age and high proficiency in English. As described in the Introduction and References therein, younger Estonians are more fluent and more exposed to English. They also are more

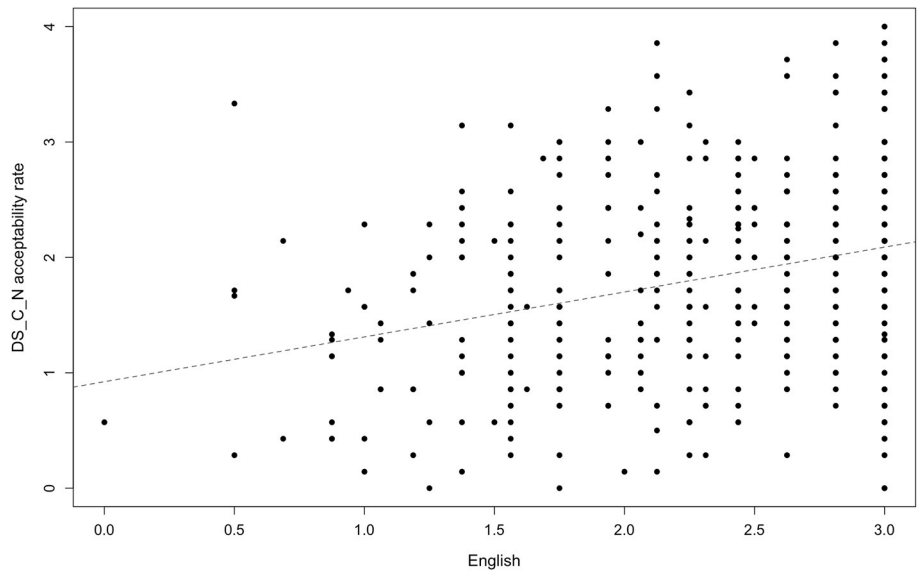

FIGURE 3 | Acceptability rate of *basic*-type adjectives in the non-agreement condition by exposure to and proficiency in English.

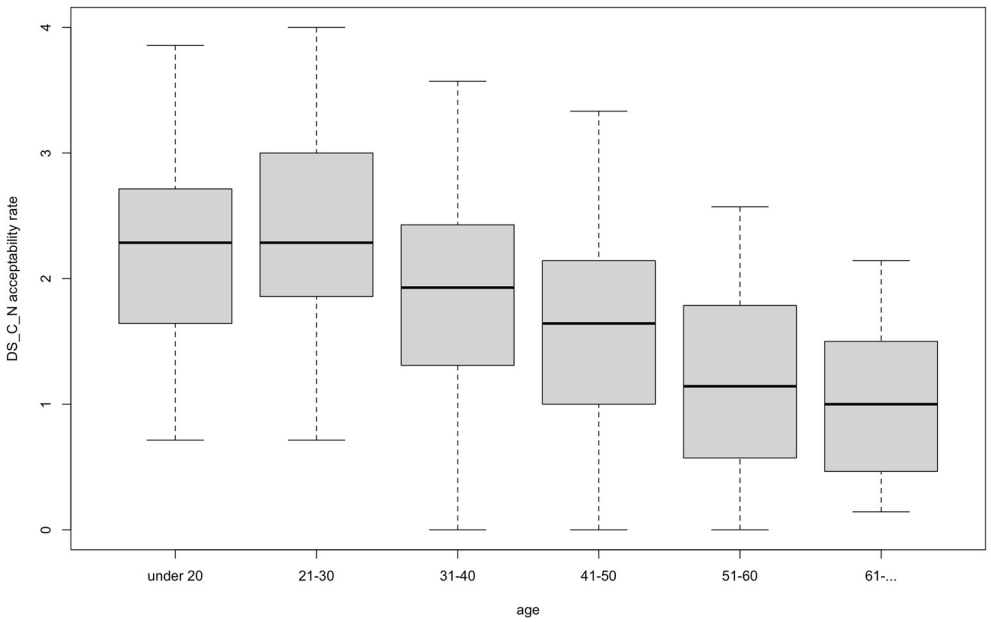

FIGURE 4 | Acceptability rate of *basic*-type adjectives in the non-agreement condition by age.

likely to use the language regularly and to produce bilingual speech, such as blogs and vlogs, as discussed in Kask (2019) and other similar studies demonstrating that bilingual users are typically under 30 years of age. Naturally, we cannot exclude puristic orientation or personal language separation ideals among younger individuals; yet, as expected, high proficiency in English and young age of the participants correlate positively with higher acceptability of all categories of adjectives. This correlation, however, does not justify using these factors interchangeably as young age and high exposure to and fluency in English have a similar but not identical effect.

The second recurrent observation is the relationship between experience of living abroad and fluency in English. Most participants with this experience were also fluent in English, but living abroad was not a prerequisite for English proficiency for young Estonians. Language proficiency and residence in an English-speaking country are often connected but these are

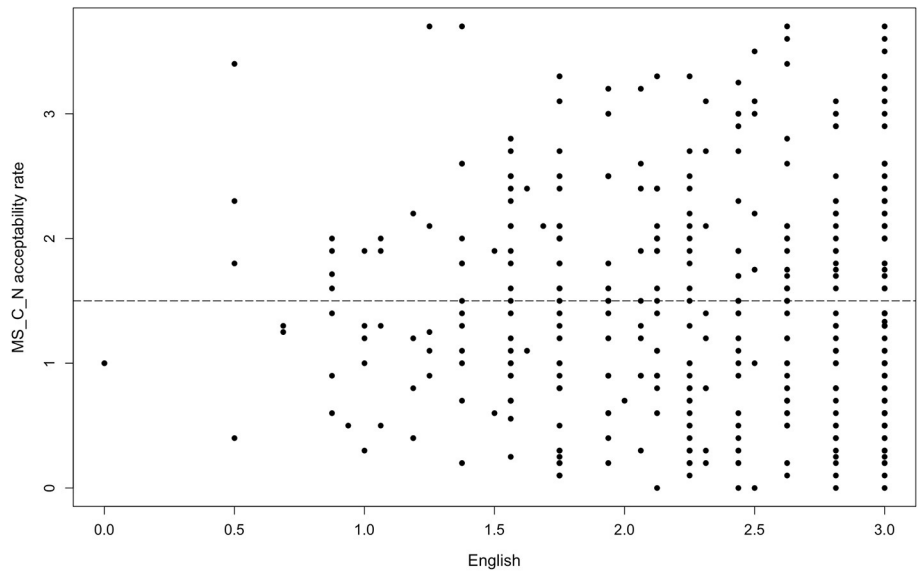

FIGURE 5 | Acceptability rate of *deep*-type in the non-agreement condition by proficiency in English.

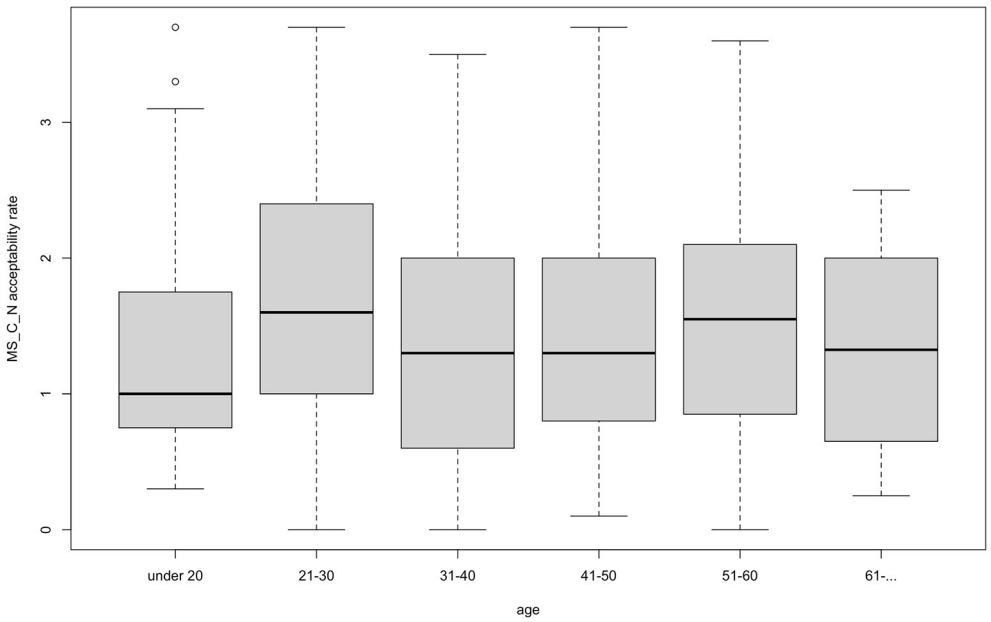

FIGURE 6 | Acceptability rate of *deep*-type in the non-agreement condition by age.

not the same: one may achieve a high proficiency in English and use the language frequently without leaving Estonia.

Another sociodemographic factor was residency in Tallinn vs. rural areas. Being a resident of Tallinn was associated with a higher acceptability rate of real sentences when compared to residents in the rural areas. Tallinn residents demonstrated significantly higher acceptability ratings for the non-agreeing disyllabic adjectives ending with a consonant (*basic*) than both rural residents and those currently residing abroad. Tallinn residents also had significantly higher acceptability rates than rural residents for the non-agreeing disyllabic adjectives ending with [i] (*fancy*) and for the agreeing monosyllabic adjectives ending with consonant (*deep*). While the pattern was not statistically significant for all adjective types, we conclude that Tallinn residency is generally positively associated with higher acceptability rates when compared to more rural areas.

The birthplace of respondents also had a relationship with acceptability rates of some adjective types: the agreeing disyllabic adjectives ending with a [i] (*fancy*) and the agreeing monosyllabic type ending with consonant (*deep*). For the latter, respondents born in Tallinn or Tartu rated this adjective type higher than respondents born in smaller Estonian towns. We may have not been able to detect other significant patterns due to low counts in some birthplace categories, but the direction is in line with what we saw with residency: respondents in bigger cities tend to have higher acceptability rates for various types of adjectives.

Gender had some effect on perception of real sentences: participants who identified as males showed slightly lower acceptability of real examples. We have no reasons to associate this pattern with structural reasons and instead interpret this as a reflection of social norms that participants may have oriented to as they completed the tests. The real sentences came from women-written and women-oriented blogs and vlogs on fashion and beauty. We consider these sentences gender-sensitive because the male respondents may find some topics less conventional for their gendered identities. The fact that we found no significant differences between gender categories in the constructed sentences suggests that the overall sentence design was well-balanced and we were able to avoid some possible gender-specific effects in acceptability rates. Gender also affected acceptability of the monosyllabic obstruent ending adjectives (*flat*): participants identifying as females had higher acceptability rates for this adjective type both in the agreement and non-agreement conditions. Currently, we are unable to explain it and a larger corpus study would shed more light on this finding.

Real vs. Constructed Sentences

While the overall experiment design removed some biases, such as gender bias, there was a minor preference for real sentences. The results showed that real sentences were rated slightly higher than constructed sentences: an average real sentence was rated around 1.8/4, an average constructed was rated around 1.5/4. The difference is minor but statistically significant and therefore something that should be considered. The results may be an artifact of lexical frequency, i.e., certain English adjectives are more frequent in bilingual speech and more conventionalized than others. At present, however, there is no way to check this, as the only existing corpus of English-Estonian bilingual communication has been created by the present authors and is about 300,000 words from fashion and beauty blogs and vlogs, so it represents a rather narrow segment of reality. The respondents may be more familiar with the adjectives used in real sentences, which leads to higher acceptability rates. In addition to that, male respondents assessed constructed sentences somewhat higher, which shifted the results for the entire dataset (see Sociodemographic Basics section for discussion on gendered blogs).

Category of Adjectives and Acceptability Rate

Increased proficiency in and exposure to English were strongly associated with higher acceptability rate of adjectives in all categories: real and constructed sentences as well as agreeing and non-agreeing adjectives. The relationship was the opposite for age: older speakers had a lower acceptability rate whereas younger speakers demonstrated a higher acceptability rate. The pattern is aligned with a general trend discussed in Sociodemographic Basics section: younger Estonians tend to be more fluent in English, and in our experiment both these factors are strongly linked to increased adjective acceptability.

Adjective Type and Acceptability Rate

We noticed the same robust pattern, dependent on the age and proficiency in/exposure to English, for all types except the *deep*-type (monosyllabic adjectives ending in a consonant) in the non-agreement condition. Neither age nor English as factors were able to account for varying acceptability rates of this specific adjective type. The factor of birthplace and residence, however, had some influence: respondents in or from larger cities had a higher acceptability rate than those from smaller towns.

Multivariate Analysis

Given the fact that our corpus consisted of data produced by 401 respondents, we had to narrow down the selection of random effects in the multivariate analysis. We selected our fixed factors based on predictors identified in the bivariate tests (age and English) and added two best random factors (residence and gender). The results show that both age and proficiency are statistically significant predictors of the outcome, with opposite effect: exposure to and proficiency in English positively affect the acceptability outcomes, whereas age slightly decreases acceptability even in respondents with advanced levels of English. These findings corroborate our bivariate tests and confirm our main hypothesis regarding age and English.

What Cannot Be Explained

Currently, we are unable to explain the exceptional behavior of the above-mentioned *deep*-type. Only the birthplace and residency factors had some impact on its acceptability in the non-agreement condition. Other than that, the acceptability did not depend either on the categories (agreement/non-agreement, real/constructed) or on sociodemographic factors.

Conclusions

Thus, our research questions are answered in the following way. The first research question, whether there is a difference in grammaticality judgement depending on proficiency in and exposure to English, is answered positively. Proficiency in and exposure to English correlated with higher acceptability of all sentences. Our hypothesis was expanded to include age because proficiency in English is higher among younger speakers.

The answer to the second research question, whether there is a difference in perception of real and constructed examples, is that the difference is only minor. Acceptability of real sentences was somewhat lower among male respondents; this has probably to do with the topic of blogs and vlogs from where the real sentences were retrieved rather than with structural factors. At the same time, overall acceptability of real sentences was slightly higher. A possible explanation may be higher frequency of some adjectives that appeared in the real sentences; at the moment

we have no possibility to check this but this is something to be considered in future. In general, the idea to propose both real and constructed sentences to participants as suggested by Verschik (2006) appears reasonable: grammaticality judgement tasks should contain some real examples, otherwise we just obtain a picture of the respondents' opinions of what is possible but cannot relate it to the real speech.

The third research question was, whether there is a difference in perception of the group of English adjectives that fit into Estonian declension system but do not have Estonian inflections in our corpus. We do not have a clear straightforward answer here, as we were not able to identify any clear patterns. In fact, some were the opposite of what one would predict based on compatibility with Estonian declension system, like *deep*, so, apparently structural factors have a limited effect and we should consider other factors, like sociodemographics, and possibly personal preferences.

In addition, multilingual cognition (multicompetence) as well as individual preferences may also play a role. Multilinguals have more linguistic resources and more developed metalinguistic awareness than a monolingual do, and multilinguals can combine elements of both grammars. Hence, a multilingual may not need the same kind of mechanisms of morphosyntactic integration, as was noted by Leisiö (2001). Earlier studies (Zabrodskaja, 2009; Zabrodskaja and Verschik, 2014) demonstrated that multilinguals do not always need to add inflectional morphology of the base language for understanding. Addition of inflectional morphology may be a matter of personal preference and depend on individual proclivity.

The results suggest that "native speakers" do not constitute a homogenous group. From the point of view of sociolinguistics, this is nothing new because different social backgrounds, social networks and linguistic environments affect language use. Still, individuals should not be treated as typical representatives of a socio-demographic group, and their linguistic behavior and metalinguistic awareness may differ. Also from the point of view of contact linguistics, changes in L1 may be perceived differently by bi- and monolingual speakers (Dogruöz and Backus, 2009). Based on the results yielded in our experiment, we believe that even bilingual speakers may differ in their grammaticality judgment. In more general terms, this means that everyone is a "native speaker" of their own idiolect, and because these idiolects have a lot in common, an illusion of a clearly definable ideal native speaker appears.

We identified several directions that would open up avenues for future research aimed at uncovering additional patterns in bilingual constructions. A larger corpus study would allow for a more detailed analysis of the frequency impact on acceptability rates. Another recommendation for an experimental design would be to look more closely into adjective constructions the acceptability ratings of which showed no correlation with the factors tested in this experiment. Finally, while this study focused on English adjectives in Estonian, it would be useful to explore other emergent bilingual constructions.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because de-identified datasets available by request. Requests to access the datasets should be directed to Anna Verschik, annave@tlu.ee.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The first observation of differential behavior of English adjectives in bilingual phrases was made by AV. Based on this, in 2019, HK conducted an empirical research, demonstrating that this was indeed the case. She identified adjective types and tendencies for (non-)agreement for each type in her data. The idea for the current study was outlined by both HK and AV, based on the method used in an earlier research by AV. The experiment was designed by all three authors. DB oversaw coding work, conducted statistical analyses, and interpreted the results. All authors contributed to the article and approved the submitted version.

FUNDING

The research was funded by the Ministry of Education and Research of the Republic of Estonia, Grant EKKD33 Data and corpora of Estonian children and youth multilingual communication.

ACKNOWLEDGMENTS

We are grateful to the members of our research team (Reili Argus, Geidi Kilp, Andra Kütt, Olga Gerassimenko, and Kadri Vider) as well as to Piret Braid and Giovanni Rossi.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.735232/full#supplementary-material

REFERENCES

Ariste, P. (1981). *Keelekontaktid*. Eesti keele kontakte teiste keeltega. [Language contacts. Estonian's contacts with other languages.] Tallinn: Valgus.

Backus, A. (2012). *A Usage-Based Approach to Borrowability*. Tilburg Papers in Culture Studies. Tillburg: Tilburg University. Available online at: https://pure.uvt.nl/ws/portalfiles/portal/30357235/TPCS_27_Backus.pdf (accessed September 24, 2021).

Backus, A. (2014). "Towards a usage-based account of language change: implications of contact linguistics for linguistic theory," in *Questioning Language Contact: Limits of Contact, Contact at its Limits,* ed. R. Nicolaï (Leiden and Boston: Brill), 91–118. doi: 10.1163/9789004279056_005

Bates, D. M., Maechler, M., and Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999–999990.

Cook, V. (2016). "Premises of multi-competence," in *The Cambridge Handbook of Linguistic Multi-Competence*, ed. V. Cook and L. Wei (Cambridge: Cambridge University Press), 1–25. doi: 10.1017/CBO9781107425965.001

Cook, V. (ed.) (2002). *Portraits of L2 Users.* Clevedon: Multilingual Matters. doi: 10.21832/9781853595851

Dewaele, J.-M. (2017). "Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker'". *Appl. Ling.* 39, 326–340. doi: 10.1093/applin/amw055

Dogruöz, S., and Backus, A. (2009). Innovative constructions in Dutch Turkish: an assessment of ongoing contact-induced language change. *Biling. Lang. Cogn.* 12, 41–63. doi: 10.1017/S1366728908003441

Eesti Õigekeelsussõnaraamat ÕS (2018). *Ed. Maire Raadik*. Tallinn: Eesti Keele Sihtasutus. Available online at: https://www.eki.ee/dict/qs/ (accessed September 24, 2021).

Erelt, M. (2007). "Preface," in *Estonian Language*, ed. M. Erelt (Tallinn: Teaduste Akadeemia Kirjastus), 7–8.

Eurostat (2017). *Foreign Language Learning.* Available online at: https://ec.europa.eu/eurostat/documents/2995521/7879483/3-23022017-AP-EN.pdf/80715559-72ba-4c19-b341-7ddb42dd61a6?t=1487690569000 (accessed September 24, 2021).

Gnevsheva, K. (2017). "Within-speaker variation in passing for a native speaker." *Int. J. Biling.* 21, 213–227. doi: 10.1177/1367006915616197

Kask, H. (2019). To agree or not to agree? English adjectives in Estonian-English bilingual blogs and vlogs. *J. Eston. Finno-Ugric Ling.* 10, 85–123. doi: 10.12697/jeful.2019.10.2.06

Kask, H. (2021). *English-Estonian Code-Copying in Estonian Blogs and Vlogs*. Doctoral thesis. Tallinn: Tallinn University Press.

Koreinik, K., and Tender, T. (2013). "Eesti keeltest rahvaloendustel [About languages in Estonia in Population Censuses]," in *Emakeele Seltsi aastaraamat [The Year Book of Mother Tongue Society]*, Vol. 59, eds. M. Erelt and S. Mäearu (Tallinn: Eesti Teaduste Akadeemia), 77–102. doi: 10.3176/esa59.04

Kruusvall, J. (2015). "Keelteoskus ja keelte praktiline kasutamine [Language Skills and the Practical Use of Languages]," in *Eesti ühiskonna lõimumismonitooring 2015 [Monitoring of Integration in Estonian Society in 2015]*, ed. K. Kallas, R. Vetik, J. Kruusvall, E. Saar, J. Helemäe, L. Kirss, C. Leppik, K. Seppel, K. Kivistik, and P. Ubakivi-Hadachi (Tartu: Haridus- ja Teadusministeerium), 72–87.

Leisiö, L. (2001). Integration and gender assignment in Finland Russian. *SKY J. Ling.* 14, 87–119. Available online at: http://www.linguistics.fi/julkaisut/sky2001.shtml (accessed September 24, 2021).

Murahata, G., Murahata, Y., and Cook, V. J. (2016). "Research questions and methodology of multi-competence," in *The Cambridge Handbook of Linguistic Multi-Competence,* ed. Li Wei and V. J. Cook (Cambridge: CUP), 26–49. doi: 10.1017/CBO9781107425965.002

Muysken, P. (2000). *Bilingual Speech: A Typology of Code-Mixing, Vol. 11*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1997). *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.

Piller, I. (2002). Passing for a native speaker: identity and success in second language learning. *J. Socioling.* 6, 179–206. doi: 10.1111/1467-9481.00184

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org

Rampton, M. B. H. (1990). Displacing the 'native speaker': expertise, affiliation and inheritance. *ELT J.* 44, 97–101. doi: 10.1093/eltj/44.2.97

Skutnabb-Kangas, T. (1984). *Bilingualism or Not?* Clevedon: Multilingual Matters.

Tammemägi, A., and Ehala, M. (2012). Koolinoorte keelehoiakud 2011. aastal [Language Attitudes of Estonian School Students in 2011]. *Keel Kirjandus* 4, 241–260.

Thomason, S., and Kaufman, T. (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley, Los Angeles, Oxford: University of California Press. doi: 10.1525/9780520912793

Verschik, A. (2006). Convergence in Estonia's Russian: directional vs. static vs. separative verb. *Int. J. Biling.* 10, 383–404. doi: 10.1177/13670069060100040101

Verschik, A., and Kask, H. (2019). "Estonian–English code alternation in fashion blogs: structure, norms and meaning," in *Multilingualism in the Baltic States. Societal Discourses, Language Policies and Contact Phenomena,* ed. S. Lazdina and H. F. Marten (London: Palgrave Macmillan UK) doi: 10.1057/978-1-137-56914-1_10, 307–335.

Zabrodskaja, A. (2009). Evaluating the Matrix Language Frame model on the basis of a Russian-Estonian codeswitching corpus. *Int. J. Biling.* 13, 357–377. doi: 10.1177/1367006908346629

Zabrodskaja, A., and Verschik,. A. (2014). Morphology of Estonian items at the interface of Russian-Estonian language contact data. *Socioling. Stud.* 8, 449–474. doi: 10.1558/sols.v8i3.25505

Zenner, E., Backus, A., and Winter-Froemel, E. (2019). "Introduction," in *Cognitive Contact Linguistics. Placing Usage, Meaning and Mind at the Core of Contact-Induced Variation and Change*, ed. E. Zenner, A. Backus, E. Winter-Froemel (Berlin: Mouton De Gruyter), 1–20. doi: 10.1515/9783110619430-001

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 October 2021
doi: 10.3389/fpsyg.2021.717639

The Native Speaker in Italian-Dialects Bilingualism: Insights From the Acquisition of Vicentino by Preschool Children

Emanuela Sanfelici[1]* *and Maja Roch*[2]

[1] Department of Linguistic and Literary Studies, University of Padua, Padua, Italy, [2] Department of Psychology, University of Padua, Padua, Italy

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez,
University of the Balearic Islands,
Spain

Reviewed by:
Tanja Kupisch,
University of Konstanz, Germany
Maria Garraffa,
University of East Anglia,
United Kingdom

***Correspondence:**
Emanuela Sanfelici
emanuela.sanfelici@unipd.it

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

Received: 31 May 2021
Accepted: 31 August 2021
Published: 15 October 2021

Citation:
Sanfelici E and Roch M (2021) The Native Speaker in Italian-Dialects Bilingualism: Insights From the Acquisition of Vicentino by Preschool Children. Front. Psychol. 12:717639. doi: 10.3389/fpsyg.2021.717639

This paper investigates the bilingualism originating from the native competence of a standard language (Italian) and a vernacular non-standardized local dialect (henceforth, bilectalism). We report results on the comprehension and production of narrative stories by 44 3- to 5-year-old typically developing children exposed to both Italian and Vicentino from birth. Our findings show that all children produced and comprehended Italian. As for the dialect, children can comprehend Vicentino, despite not producing any dialectal element. The study further revealed an implicational scale in dialectal competence: if a child exhibits some productions with dialectal syntax, s/he also produces dialects at the phonological, morphological, and lexical levels. These findings are in line with the dialectological studies on adult speakers: dialectal competence should be arranged along a fine-grained continuum and the dialectal speaker should be considered as a multi-factorial notion. Our study extends this observation to children's dialectal acquisition.

Keywords: bilingualism, Italian dialect, comprehension, production, preschool children

INTRODUCTION

Recent linguistic and psychological literature has questioned the standard definition of the native speaker suggesting that nativeness is an articulated, composite, and dynamic state which may result from the interaction between linguistic cognition and extra-linguistic factors, e.g., literacy (Davies, 2003, 2013; Sorace, 2003; Hulstijn, 2015; a.m.o.). Whereas much attention has been paid to second language acquisition, bilingualism in two majority language contexts, and heritage language acquisition, little is known about the notion of nativeness and, more generally, about the acquisition modes in bilectal contexts (Lauchlan et al., 2012; Vangsnes et al., 2015; Antoniou et al., 2016; Leivada et al., 2017) and even less about the acquisition of Italian-dialects bilectalism (with the exception of Sardinian, Garraffa et al., 2015, 2017; Klaschik and Kupisch, 2016; Kupisch and Klaschik, 2017).

Italian-dialects bilectalism provides an ideal ground for a deeper understanding of nativeness. Indeed, sociolinguistic and dialectological studies on Italian dialects have demonstrated that, besides sharing some features with other types of bilingualism, Italian bilectalism exhibits at least three outstanding properties. First, whereas the whole community usually comprehends both Italian and the local dialect and speaks Italian, only a portion of the community can also speak the local dialect (Mioni and Trumper, 1977; Mioni, 1979, 2007; Marcato and Ursini, 1998; Loporcaro, 2013; Garraffa et al., 2017; a.o.). The ISTAT (2015) survey revealed that more than half of the Italian population from age six can only speak Italian (55.1%). Both dialect and Italian are used in everyday life by 25.2% of Italians, while only 12.2% actively speak the local dialect.[1] The low proportion of dialectal speakers may affect the quantity of dialectal input children are exposed to. Children may be more frequently exposed to Italian since Italian is likely to be used more frequently, e.g., in school and the media, and by many different native speakers. Second, the local dialects and Italian are used in different sociolinguistic domains (Mioni, 1979, 2007; Berruto, 2006; a.m.o.). Dialects are often bound to specific and limited contexts, like home language. While the majority of Italians speak exclusively Italian when interacting with strangers (79.5%) and friends (49.6%), Italians speak both languages (32%) or only dialect (14.1%) at home, according to the ISTAT (2015) survey. Since hearing a language from several different speakers is more supportive of language development than hearing a language for the same number of hours from fewer speakers (Houston and Jusczyk, 2000; Hoff, 2015), the lack of multiple speakers may lead to a less rich and less varied dialectal input. Third, dialects often lack a written standard: with very few exceptions (e.g., Friulian), no children's books and no media activities are available in dialect. In addition, minority languages and dialects usually lack official recognition, which implies less or null support at the educational level. It is usually the case that the local dialect is not taught in school. Since all these extra-linguistic properties plausibly influence input consistency, bilectal acquisition may differ from the acquisition in standard bilingual contexts.

The dialectological literature has also demonstrated that dialectal competence is arranged along a fine-grained continuum and that dialectal speakers do not have a fixed position on this scale along their life span (Mioni, 1979; Berruto, 2005, 2006; Benincà and Poletto, 2006; Cerruti, 2011; a.o.). Various studies on adults and school-aged speakers have shown that, while only a few students attending middle schools declared they were competent in their local dialect, the number of dialectal speakers increased with age (Berruto, 2005; Marcato, 2005; Berruto, 2007; Tessarolo and Bordon, 2015). While adults' and adolescents' dialectal competence has been quite extensively investigated, little information is available on the early stages of dialectal acquisition (Cardinaletti, 2013; Bonifacio, 2014; Garraffa et al., 2015; Klaschik and Kupisch, 2016; Kupisch and Klaschik, 2017; Covazzi, 2019). Some studies have investigated children's dialectal production (Cardinaletti, 2013; Bonifacio, 2014; Klaschik and Kupisch, 2016; Kupisch and Klaschik, 2017; Covazzi, 2019), whereas others have tested comprehension (Garraffa et al., 2015). All these acquisition studies have demonstrated that, although Italian is the dominant language, at least some children are competent in the local dialect. Children's dialectal competence varies across children and across the linguistic phenomena investigated in the studies. Furthermore, according to some scholars, children's dialectal production depends on the quantity of dialectal input children have been exposed to Klaschik and Kupisch (2016), Kupisch and Klaschik (2017); but see Covazzi (2019). We lack a study that assesses both comprehension, production, and quantity of input in bilectal children.

Building on the previous acquisition studies, we investigate both production and comprehension of Vicentino, a Venetan dialect, by preschool children from age 3 to 5. Through a parental questionnaire, we measure the quantity of dialectal input to verify to which extent children's production and comprehension correlate with language exposure. The broad aim of our study is to assess the nature and the status of the bilectal preschool speaker, by asking whether preschool children comprehend and produce any dialect.

The structure of the paper is as follows. In Section "Previous studies on early bilectal acquisition" we summarize previous studies on early bilectal acquisition. In Section "The Vicentino dialect" we outline some notes on the sociolinguistic status of Vicentino and delineate those grammatical properties differentiating Vicentino from Italian, that become relevant in the analysis of children's productions. Section "Current study" presents the study, while Section "Results" reports the results. Finally, in Section "Discussion" we discuss our findings and conclude the paper.

PREVIOUS STUDIES ON EARLY BILECTAL ACQUISITION

In the last decade, scholars have started investigating early bilectal acquisition, asking how bilectal children perform in grammatical and cognitive tasks and how they acquire various dialectal grammatical properties. Some studies have demonstrated that bilectal children's performance is not distinct from that of monolingual children, and, when differences emerge, bilectal children outperform monolinguals (Cardinaletti, 2013; Garraffa et al., 2015). Garraffa et al. (2015) investigated receptive Italian grammatical competence and general cognitive abilities in bilectal Italian-Sardinian children attending primary school. Their results showed that the performance of Sardinian–Italian bilectal children was generally indistinguishable from that of monolingual Italian children, in terms of both Italian language skills and cognitive abilities. Where differences were detected, these were mostly in favor of bilectal children. A similar conclusion was reached by Cardinaletti (2013), who analyzed the emergence of clitics in the spontaneous productions of one bilectal Italian–Venetan child aged 2–3 years, one monolingual Venetan child, and one age-matched Italian monolingual child. In the bilectal child, object clitics emerged roughly at the same

[1] The remaining percentage refers to the use of another (foreign) language.

time in Italian and Venetan, and omissions stopped at the same time in both languages. Conversely, in the two monolingual children object clitics were omitted at a much higher rate and for a longer period than in the bilectal child. The author concluded that converging evidence from two close languages, Italian and Venetan, speeds up the acquisition of object clitics. Conversely, Bonifacio (2014) did not find a similar advantage. The author investigated the acquisition of the copula in the spontaneous productions of two bilectal Italian-Venetan children (Age 19–27 months and 24–36 months) and one Italian monolingual child aged 22–28 months. Although the three children exhibited a very similar developmental path in copula omission and production as well as the agreement between the copula and the subject, they differed in the productions of articles in copular contexts. While the Italian monolingual child produced articles in a target-like fashion, the two bilectal children were not fully competent, producing very few articles.

Only three studies we are aware of provide information concerning the quantity of dialectal input children are exposed to and children's dialectal productions (Klaschik and Kupisch, 2016; Kupisch and Klaschik, 2017; Covazzi, 2019). Klaschik and Kupisch (2016) tested the interpretation of subjects in Italian and Venetian by 20 Italian–Venetian children aged 7–12 years. Bilectal speakers were reported to accept more overt subjects when the discourse topic was maintained compared to monolingual Italian speakers. This result suggests that there was an influence from Venetian, where overt subjects are mandatory in some contexts. In addition, those children who were reported to use dialect in everyday life comparatively more often outperformed children who were reported to be infrequent dialect users. The authors concluded that exposure to a dialect does not negatively affect the use of the standard language. Kupisch and Klaschik (2017) study dialectal influence and gender marking in 25 bilectal Italian–Venetan children aged 5–11 years. Participants were tested with an elicited production task in both Italian and Venetan. Children followed the gender assignment rules of Italian and Venetan. Interestingly, no dialectal influence was detected in the Italian experiment: children produced only Italian DPs. Conversely, in the Venetan experiment, the production was much more varied: children produced Italian, Venetan, and mixed DPs. The degree of dialectal use was suggested to be dependent on the quantity of dialectal input children may have been exposed to. A slightly different conclusion was reached by Covazzi (2019), investigating the production of relative clauses by 23 preschool bilingual Italian–Friulian children aged 4–6 years. In line with much cross-linguistic research, she found that the production of subject relative clauses was more accurate than that of object relatives. Although Italian was the dominant language and children's production was essentially Italian, the author reported that specific influences of Friulian on Italian were indeed present in children's productions. However, no correlation was found between children's Friulian productions and the quantity of Friulian input children were reported to receive in the parental questionnaire.

Taking the findings together, we conclude that, although children may be competent in the dialect, the degree of children's dialectal production and comprehension seems to depend on the linguistic phenomenon investigated and the quantity of dialectal input. Our study adds to these previous acquisition works by testing both comprehension and production and by including a measure of dialectal input. Before illustrating the study, we outline some notes on the sociolinguistic status and grammatical properties of Vicentino.

THE VICENTINO DIALECT

Venetan is a Western Romance language. Similar to other varieties, Venetan has developed independently from Latin. Thus, it is not a dialect of Italian from a genealogical perspective. As Zamboni (1974, 1979) proposed, Venetan comprises different language groups: *the Venetian group* (Venezia), *the central group* (Padova, Vicenza, Polesine), *the North Venetan group* (also labeled trevigiano-feltrino-bellunese), the *outer Venetan* spoken in the Trento province, and the varieties spoken in Friuli Venezia-Giulia, Istria and Dalmazia (Zamboni, 1974, 1989; Cortelazzo, 1981; Loporcaro, 2013).

Vicentino is a Venetan dialect, spoken in the Vicenza province located in the Veneto region (Pellegrini, 1977). In 2007, the Veneto regional administration issued a law (L.R. 13.4.2007, n. 8) promoting and financially supporting initiatives to study and preserve the use of the local Venetan dialects. Thanks to this law and the regional economic growth between the '80s and '90s, Venetan dialects, including Vicentino, have recently experienced an increase in their prestige, especially among young people aged 25–34 years (Santipolo and Tucciarone, 2006). According to the ISTAT (2015) survey carried out more than half of the Veneto population speak the local dialects at home. One third of the Veneto population exclusively speak a Venetan dialect at home (30.6%), one third use both Italian and a Venetan dialect (31.4%), while the remaining Veneto population exclusively use Italian (28.5%). A similar observation holds for the use of dialect when speakers interact with friends: 28.7% of the population exclusively use a Venetan dialect, 33.6% use both Italian and a Venetan dialect, while 30.6% only speak Italian. With strangers, the use of dialect decreases: people speak either exclusively Italian (65.6%) or use both Italian and dialect (23.1%), whereas only 8.7% only speak the local dialect.

Italian and Vicentino differ with respect to several phonological, morphological, syntactic, and lexical properties. In this section, we provide a summary of the properties differentiating Vicentino from Italian: the list is by no means exhaustive but contains only those features relevant to the results in Section "Results."

Some grammatical features of Vicentino are shared by other Venetan dialects, whereas others are peculiar to this group. Unless otherwise specified, we do not distinguish when the features are common to the Venetan dialects or only specific to the central Venetan ones, to the extent that they differ from those of Italian.

Among the phonological dialectal features, we find voicing of degeminated plosive in intervocalic contexts: degeminated voiceless plosive consonants become voiced in intervocalic contexts, [rɔ:da] for *ruota* 'wheel,' [manè:go] for *manico* 'handle' [see ex. (6) in Section "Results"]. In addition, phonological

double plosives undergo degemination in intervocalic contexts: [katì:vo] instead of the Italian [kattì:vo] 'bad' (Zamboni, 1974; Loporcaro, 2013, p. 86). Likewise, the lateral consonant/l/in intervocalic contexts undergoes degemination when geminated, [bà:la] instead of the Italian [pàlla] 'ball' (Zamboni, 1974, p. 37) [see ex. (6) in Section "Results"]. Vicentino also has the apocope of the final unstressed vowels/e/and/o/. In other Veneto dialects, i.e., Venetian, apocope of the final unstressed vowel/e/occurs after degeminated/n, l, r/, while/o/is dropped only when following/n/: [kà:ŋ], [sa:l], [saòr] instead of the Italian [kà:ne] 'dog,' [sà:le] 'salt' and [sapò:re] 'taste'; [feŋ], [pjeŋ] instead of [fjeno] 'hay' and [pjeno] 'full.' In Vicentino, apocope of/e,o/is more restricted: it can only occur when following/n/(see Loporcaro, 2013, p. 86, 105). Hence, while in Vicentino 'dog' is [kà:ŋ], the word for 'salt' is [sà:le], without the apocope of/e/. Notice that the final/n/undergoes velarization after apocope [see ex. (4–5) in Section "Results"].

Among the morphological features, we quote the unique form of the masculine singular definite article which is a weak form and is *el*, instead of the *lo/il* of Italian (Marcato and Ursini, 1998, p. 86).

A pervasive feature of the Vicentino dialect is the productive use of prefixes attached to denominal and deadjectival verbs to encode various aspectual values (Marcato and Ursini, 1998, p. 230). A case in point is represented by the prefix *in-*: *in-rabiarse* (It. *arrabbiarsi*) 'to get angry,' *in-tardigare* (It. *tardare*) 'to be late,' *in-marsire* (It. *marcire*) 'to rot.' Another instance is the use of the prefix *s-* instead of the Italian *in-*, as in *s-cumissiare* (It. *in-cominciare*) 'to begin, start.'

As for the morpho-syntax, Vicentino productively uses particle verbs. Some examples are *dire sù* 'lit. to speak up; to reproach,' *saltare sù* 'lit. to jump up; to attack,' *torre sù* 'lit. to take up; to pick up,' *molare zò* 'lit. to release down; to release.'

While in Italian the progressive imperfective aspect is encoded by the construction *stare* 'to stay' plus gerund, Vicentino uses the construction *essere/essare drio* 'lit. to be behind' plus bare infinitive as exemplified in (1) [see ex. (6) in Section "Results"].[2]

(1) (a) Italian
Sto *mangiando*
Stay.1sg eating

(b) Vicentino
So *drio* *magnare*
be.1sg behind eat
'I am eating.'

Another difference between Italian and Vicentino lies in the pro-drop parameter value. As is well-known, Italian is a pro-drop language (see Rizzi, 1982). Conversely, Vicentino does not generally allow a pro subject: subject clitics for III singular and plural person must be lexicalized when the predicate is finite (see Benincà and Vanelli, 1982, p. 41; Marcato and Ursini, 1998; we refer the reader to Benincà, 1994; Poletto, 1996 for a more precise picture on subject clitics in Vicentino). With existential predicates, Vicentino exhibits the locative clitic *ghe* (Benincà and Vanelli, 1982), as in *ghe jera un can* 'there was a dog.'

Another difference between Italian and Vicentino syntax regards auxiliary selection. In finite contexts, in Vicentino causative transitive, impersonal, antipassive, and reflexive predicates allow both the auxiliary *ésar* 'to be' and *gavere* 'to have,' with a general preference for *gavere* (Gagarina et al., 2012; Benincà, 1994; a.o.). Conversely, in Italian, these predicates only select the auxiliary 'to be.' Compare the Vicentino examples in (a) with the Italian counterpart in (b) (see ex. (7) in Section "Results").

(2) (a) *I* *tosi* *gà* *cambià*
the boys have.3pl changed

(b) *I* *ragazzi* *sono* *cambiati*
the boys be.3pl changed
'The boys changed.'

(3) (a) *El* *baonsin se* *ga* *incastrà* *sora* *l'* *albaro*
the balloon cl.refl have.3sg wedged on the tree

(b) *Il* *palloncino* *si* *è* *incastrato* *su-ll'* *albero*
the balloon cl.refl be.3sg wedged on-the tree

Finally, we quote the presence of the doubly filled comp in various subordinating conjunctions, which is absent in Italian: the complementizer *che* follows the wh-phrase *come* 'how,' i.e. *come che* 'how,' and the adverbial *péna* 'soon,' i.e. *péna che* 'as soon as' (Marcato and Ursini, 1998). We further notice that Vicentino temporal subordinate clauses are introduced by the lexical item *có* instead of the corresponding Italian *quando* 'when' (see ex. (8) in Section "Results").

Although this linguistic profile is not exhaustive, it allows us to detect clear phenomena in which the Vicentino dialect and Italian diverge and thus to characterize the linguistic productions by establishing the type and the degree of dialectal features children produce.

CURRENT STUDY

We investigated the comprehension and the production of narrative stories in bilectal preschool children who have been exposed to Italian and Vicentino from birth. Participants were tested with the picture-supported task "Multilingual Assessment Instrument for Narratives" (MAIN, Gagarina et al., 2012, 2015; Gagarina and Lindgren, 2020) in two experimental sessions, one in Italian and the other in Vicentino. In addition, parental questionnaires were collected to obtain information concerning the family socioeconomic status as well as children's language exposure.

Since no studies we are aware of have tested both comprehension and production in this bilectal context, our broad research question was to determine preschool children's dialectal competence in both production and comprehension. Four research questions were formulated.

[2] Abbreviations in examples follow Leipzig's glossing conventions.

(Q1) Does children's performance in narrative comprehension and production differ depending on the mode of the narrative task, i.e., retelling vs. telling?

We expect a difference in the narrative mode. We hypothesize the complexity, coherence, and accuracy of narrative stories, i.e., macrostructure, to increase when the child is presented first with a well-structured and coherent model story and then asked to retell the story. This hypothesis is based on various studies on bilinguals showing that children comprehended better and produced more structured and coherent stories with the retelling task than with the telling task (e.g., Kunnari et al., 2016; Maviş et al., 2016; Levorato and Roch, 2020).

(Q2) Are there differences in narrative abilities between children determined by age?

A difference in age is also hypothesized given previous studies which demonstrated that children's narrative abilities increase with age (e.g., Pearson, 2002; Florit et al., 2014). This should hold for the complexity, coherence, and accuracy of children's narratives (Bohnacker, 2016; Bonifacci et al., 2018).

(Q3) Does children's performance in narrative comprehension and production differ depending on the language of the experiment, i.e., Vicentino vs. Italian?

Question (3) forms the bulk of this study. If bilectalism is similar to the bilingualism found in other contexts, we expect to find no differences depending on the language of administration, especially in the macrostructure. Story macrostructure is reported to be less dependent on language abilities as compared to microstructure, as typically operationalized in number of sentences and words (Berman and Slobin, 1994; Rodina, 2016). In addition, various studies have proposed that there might be a carry-over of the particular macrostructure elements across the bilingual's two languages, even if the child's linguistic abilities in one of them are weaker (Gagarina, 2016). As in previous bilingual studies, if there is a difference depending on the language, we expect this to affect the microstructure. Given previous dialectological and sociolinguistic findings reported in Sections "Introduction" and "Previous studies on early bilectal acquisition," we hypothesize the acquisition of dialect to lag behind that of Italian.

(Q4) Does children's competence depend on the quantity of input?

We expect the quantity of input to affect children's performance. To date, studies have indicated that the quantity of language exposure is a crucial factor in language development. The quantity of language exposure has been reported to significantly influence the size of children's vocabularies (Hoff et al., 2014) as well as the morpho-syntactic development and language proficiency (Hoff, 2015; Dicataldo et al., 2020). Since microstructure measures of bilingual narratives also reflect the lexical and morpho-syntactic development of each of their languages, we hypothesize that we will detect the effect more in the microstructure than in the macrostructure.

Participants

We tested 44 3- to 5-year-old children exposed to both Italian and Vicentino dialect from birth: 7 3-year-old children (Mean Age: 3.4; Standard Deviation: 2.7 months); 15 4-year-old children (Mean Age: 4.6; Standard Deviation: 4.2 months); 22 5-year-old children (Mean Age: 5.8; Standard Deviation: 3.9 months). In addition, 10 adults (Mean Age: 25.8; Standard Deviation: 6.9 months) were also tested. Children were recruited from three kindergartens in the Vicenza province, located in Thiene, Santorso, and Schio. All participants were typically developing children: none experienced any developmental disorders as attested by the parental report.

Background information concerning the family's socioeconomic status, length of exposure to Italian and Vicentino, the quantity of the child's production in both languages, and the quantity of input the child was exposed to in both languages, was collected through the Italian version of the Questionnaire for Parents of Bilingual Children filled out by parents (PABIQ, English version: COST IS0804, 2011; Italian version: see Dicataldo and Roch, 2020; Levorato and Roch, 2020).

According to the parental questionnaire, all children have been exposed to Italian and Vicentino from birth. There were no differences in the socioeconomic status (SES) of the parents: years of education for the mother ranged from 13 to 17 with a mean of 14 years. All children's parents reported speaking both Italian and Vicentino with their partner at home. Differences emerged in the language they used to interact with their children, the quantity of Vicentino/Italian input their children were exposed to, and, in turn, the language and the quantity of Vicentino/Italian children used in their interactions with the caregivers. Parents were asked to estimate language exposure to both Italian and Vicentino by grading how often they used Italian in comparison to Vicentino when talking with the child and conversely, how often their child replied to the caregiver in either language. Up to five caregivers interacting with the same child were allowed to be listed in the questionnaire. By summing up the grades and dividing the sum for the number of caregivers reported in the questionnaire, we calculated the average amount of time caregivers spoke and children replied in each of the two languages (see **Appendix Tables A,B**). We identified four ranges for both input and output: only Italian (10), almost exclusively Italian (9.9 > 9), mainly Italian (8.9 > 5,1), and seldom Italian, i.e., mainly dialect (5 > 0). Next, we combined the four ranges of input in Italian with those of output in Italian. **Table 1** illustrates the distribution of the 44 children across the Italian input-output ranges.

Table 1 shows that most children were exposed to and produce Italian, while only few children received input and replied in Vicentino. Thirteen children received input and replied in Italian

TABLE 1 | Children's distribution according to the quantity of input and output in Italian during their interaction time with caregivers as reported in the Italian version of The Questionnaire for Parents of Bilingual Children (Italian version: Dicataldo and Roch, 2020; Levorato and Roch, 2020).

	Input in Italian			
Output in Italian	**100%**	**99 > 90%**	**89 > 51%**	**50 > 0%**
100%	13 children	6 children	13 children	1 child
99 > 90%		1 child	4 children	
89 > 51%			3 children	
50 > 0%				3 children

only. Moreover, 20 children replied in Italian, although their input was not exclusively in Italian. For the remaining children (N = 11), the quantity of Italian in the input seems to match the quantity of Italian in the output. We ran a linear regression to determine whether the quantity of input predicted the quantity of output reported in the questionnaires. Our model revealed a significant effect of the quantity of the Italian input children are exposed to on the quantity of Italian output children use in their interactions (R^2 = 0.0, RMSE = 2.1, F = 38.2, $p \leq 0.001$).

Out of the 44 children, 11 children did not comprehend the questions in the familiarizing phase and the experimental instructions in Vicentino. In the parental questionnaires, these children were reported to use exclusively (N = 9) or almost exclusively (N = 2) Italian in their interactions with the caregivers. Since we were interested in characterizing bilectal speakers, we excluded these 11 children from further analyses. Hence, the final sample of participants consisted of 33 children: 6 3-year-old children, 9 children aged 4, and 17 5-year-old children.

Materials and Procedure

Participants were tested on the production and comprehension of Italian and Vicentino dialect with the picture-supported task "Multilingual Assessment Instrument for Narratives" (MAIN, Gagarina et al., 2012, 2015; Gagarina and Bohnacker, 2022), developed within the Narrative and Discourse Working Group (WG2) of the COST Action IS0804, "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment." We adopted the Italian version of the MAIN (Gagarina et al., 2012; Levorato and Roch, 2020). The Italian MAIN was used both in the Telling and the Retelling modes: two picture-supported stories for telling (Baby Birds, Baby Goats) and two for retelling (Cat, Dog), each consisting of three episodes. There were nine comprehension questions for each story, i.e., three questions for each episode. The Italian version of the MAIN was translated into Vicentino. The testing in Italian and Vicentino was carried out by different experimenters and on different days. The testing procedure for narrative elicitation consisted of three stages, the same in each language: (1) Familiarization phase, (2) Narrative Telling (MAIN: Baby Birds/Baby Goats, counterbalanced) and comprehension questions, (3) Narrative Retelling (MAIN: Dog/Cat, counterbalanced) and comprehension questions.

Each child was tested in both Vicentino and Italian throughout four sessions with a delay of at least two days. All children were tested first in Vicentino and then in Italian. Notably, if the story of baby birds was used in Italian in the telling mode, the same child was tested with the other story, i.e., baby goats, in Vicentino. The same observation holds for the Narrative Retelling mode. The instructions, the feedback, as well as the comprehension questions, were given in each language, Vicentino in the Vicentino experiment and Italian in the Italian experiment.

All children were tested individually by a proficient speaker of Italian and Vicentino in a quiet room in the kindergarten. The experimenter and the child were seated next to each other during the telling and retelling modes. First, the child was asked several warm-up questions, e.g., 'Do you like listening to stories and fairy tales? Do you know what a story or a fairy tale always begins with/ends with?' If the child did not know the answer, the experimenter explained how stories could begin and end. The child was also prompted to tell any story s/he wanted. Then the experimenter presented the child with three envelopes and informed the child that each contained a different story. Actually, all envelopes contained the same picture story. This step was necessary to strengthen the child's belief that the experimenter was not familiar with the stories. The child was asked to choose one envelope.

In the telling mode, the child was asked to take the picture story from the envelope, look at the pictures, and tell a story without showing the pictures to the experimenter (the child was explicitly asked not to do that). This was done to ensure the 'non-shared attention' condition, as the experimenter was only the listener and the child had to narrate alone. The experimenter prompted the child gently when s/he could not begin, or when there was a long pause. After the production of the story, the experimenter asked the child some questions to assess the child's understanding of the story. In the retelling mode, when the child had chosen the envelope, the experimenter and the child viewed the pictures together. First, the experimenter told the model story to the child in a friendly manner, following the script and pointing to the pictures (see Gagarina et al., 2012). Subsequently, she asked the child to retell the story. After the retelling, the child was also asked a set of comprehension questions.

Analysis and Scoring

We analyzed both the macrostructure and the microstructure of the children's narratives. Macro-structure is the global hierarchical organization of a text and the overall coherence of the story, while microstructure pertains to the specific types of words and syntactic structures that make up the story (Berman and Slobin, 1994). We analyzed four aspects: (a) story structure in both the Telling and the Retelling; (b) comprehension in both modes; (c) syntactic measures; (d) linguistic properties of the production.

Story Structure

A maximum of 17 points could be given for story structure in both the Telling and the Retelling mode: 2 points for expressing a setting, and a total of 15 for the three episodes of each story: within each of three episodes 1 point was given for an Initiating event (max. 3 episodes * 1 = 3 points); 3 points for each element of the Goal-Attempt-Outcome sequence (max. 3 episodes * 3 = 9 points); 1 point for the Reaction/Response (max. 3 episodes * 1 = 3 points). We illustrate the scoring with **Figure 1**, which depicts the first Episode of the Baby Goats story.

The initiating event is represented by the baby goat being scared/in danger/drowning in the water or by the mother seeing the baby goat in danger. The goal is to help the baby and rescue him. The attempt is represented by the mother goat going into the water. The Reaction/Response is the mother goat/baby goat being relieved, happy, not scared anymore.

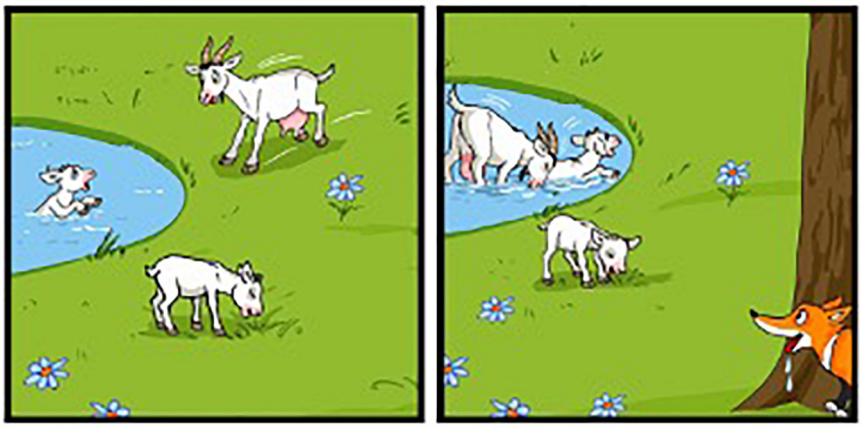

FIGURE 1 | Episode 1 of the Baby goats in the MAIN story telling (from Gagarina et al., 2012).

Story Comprehension

A maximum of 9 points could be obtained, 1 for each question answered. All the questions were cued recall questions (see Maviş et al., 2016). Three of the questions targeted the three goals (e.g., 'Why is the goat in the water?'), two questions elicited internal state terms connected either to the initiating event or reaction elements (e.g., 'How does the baby goat feel?'), and were followed by three clarification questions ('Why?'), and one question eliciting a theory of mind response (e.g., 'Imagine that the bird sees the goats. How does the bird feel?'), followed by a clarification question ('Why?').

Syntactic Measures

We coded two syntactic measures, namely the number of utterances and the mean length of utterance computed in words. Both measures were calculated independent of the language children used in their narratives. Hence, even if their utterances in the Vicentino mode were in Italian, they were included in the calculation of the number of utterances and the MLU for the Vicentino mode. This choice allowed us to detect whether the language of the instructions in both tasks and of the told story in the retelling task had an effect on children's performance, by for instance hindering their productions.

Linguistic Features

As pointed out by Kupisch and Klaschik (2017), it is not straightforward to determine whether a production qualifies as dialectal or Italian. There is indeed a high degree of linguistic overlap between Vicentino and Italian which regards the lexical, phonological, morphological, and syntactic levels: many words, morphemes, sounds, and structures are identical in the two languages. In addition, as in many bilectal communities, there are mixed codes, intermediate between Italian and the dialect, in the Veneto area. To detect the amount of Vicentino children produce we only considered those phonological, morphological, and syntactic patterns listed in Section "Previous studies on early bilectal acquisition," where the dialectal grammar diverges from Italian. Therefore, productions were classified as dialectal when they contained a feature that unambiguously qualified as Vicentino according to the dialectological studies. They were classified as Italian when they contained a feature that clearly pertained to Italian grammar. Those elements that could be classified as both Italian and Vicentino, such as the word *luganeghe* 'sausages,' were disregarded. We included phonetic properties, such as nasalization, when the audio quality allowed us to analyze children's pronunciation. Although this procedure allowed us to determine how much dialectal elements were produced, a proper quantitative analysis of children's dialectal output remains problematic since the data includes a mix of different language levels.

TABLE 2 | Descriptive statistics for Story Structure in bilingual MAIN-narratives across Age groups: mean and (standard deviation).

	Age	Italian		Vicentino	
		MAIN-Retelling	MAIN-Telling	MAIN-Retelling	MAIN-Telling
Story Structure	3	5.7 (1.2)	3.5 (0.8)	4.3 (1.3)	3.7 (0.8)
	4	6.2 (1.4)	3.9 (1)	4.7 (1.8)	3.64 (1.5)
	5	6.9 (1.9)	5.3 (1.9)	6.9 (1.7)	5.4 (1.5)

RESULTS

Story Structure

We first report results on Story Structure. **Table 2** shows the overall scores of Story Structure where children could score a maximum of 17 points in the two narrative modes of the MAIN across the two language modes.

Data were fitted to a linear mixed model. The score children obtained in the story structure was our dependent variable. Age (3, 4, 5), Mode (Telling vs. Retelling), Language (Vicentino vs.

Italian), and Quantity of Input were the fixed effect factors. Children's IDs were included as the random effects grouping factor. As for reference categories, '3-year-old' was the reference level for the Age factor, Retelling for the mode factor, and Vicentino for the language factor, and 2 for the Quantity of Input. The model revealed a significant effect of the factor Mode with the retold stories showing a higher scoring for the story structure than the told stories. In addition, a significant effect of Age was detected. No significant effect of Language and Quantity of Input was found and, likewise, no significant interaction of the three factors emerged. The results of the model are reported in **Table 3**.

We ran multiple comparisons with Tukey correction on the factor Age. The comparisons revealed that 3-year-old children did not differ from 4-year-old children (MD = –0.3; SE = 0.446; t = –0.72; p_{tukey} = 0.89) but they differed significantly from the 5-year-old ones (MD = –1.7; SE = 0.43; t = –3.984; $p_{tukey} < 0.001$). In addition, 4-year-old children significantly differed from 5-year-old ones. The analyses showed that older children, at age 5, outperformed the younger ones, both 3- and 4-year-olds.

Story Comprehension

Next, we report results on Story Comprehension. **Table 4** illustrates the overall scores of Story Comprehension where children could score a maximum of 9 points in the two narrative modes of the MAIN across the two language modes.

We fitted our data to a linear mixed model with one random intercept - participants-, one dependent variable -the score children obtained in the comprehension questions-, and three fixed effects: Age, Mode, and Language. We used the same reference levels outlined in the previous section. The model detected a significant effect of Age and Mode. No significant effect of Language and interactions were found. The results of the model are reported in **Table 5**.

Tukey *post hoc* comparisons of the main effect of Age showed that 3- and 4-year-old children significantly differed from 5-year-old peers (Age 3 vs. Age 5: MD = –1.6; SE = 0.41; t = –4; $p_{tukey} < 0.001$; Age 4 vs. Age 5: MD = –1.4; SE = 0.34; t = –4.2; $p_{tukey} < 0.001$). Conversely, 3- and 4-year-old children did not differ (MD = –0.2; SE = 0.43; t = –0.49; p_{tukey} = 0.96).

TABLE 3 | Results of the linear mixed model for the dependent variable Story Structure.

Predictors	df	*F*	*p*
Narrative-Mode	1, 27.15	33.95	< 0.001
Language-Mode	1, 18.00	1.067	0.315
Age	2, 17.86	6.705	0.007
Quantity of Input	6, 17.86	1.069	0.417
Narrative-Mode * Language-Mode	1, 35.81	4.078	0.061
Narrative-Mode * Age	2, 27.15	0.619	0.546
Language-Mode * Age	2, 18.00	0.993	0.390
Narrative-Mode * Quantity of Input	6, 27.15	0.940	0.483
Language-Mode * Quantity of Input	6, 18.00	0.788	0.591
Age * Quantity of Input	6, 17.86	0.928	0.499
Narrative-Mode * Language-Mode * Age	2, 35.81	0.438	0.649
Narrative-Mode * Language-Mode * Quantity of Input	6, 35.81	1.017	0.430
Narrative-Mode * Age * Quantity of Input	6, 27.15	1.085	0.396
Language-Mode * Age * Quantity of Input	6, 18.00	0.209	0.969
Narrative-Mode * Language-Mode * Age * Quantity of Input	6, 35.81	0.321	0.922

The model was fitted using restricted maximum likelihood. Full model summary: N = 132, REML 432.2.

Syntactic Measures

Now we report results on the Syntactic Measures, i.e., number of utterances and mean length of utterance. **Table 6** shows the overall scores of Syntactic Measures which children obtained in the two narrative modes of the MAIN across the two language modes.

We first analyzed the number of utterances children produced in the two tasks in both Italian and Dialect. We performed the same analyses described for Story Structure and Comprehension. The linear mixed model revealed a significant effect of the factor Mode, with the retelling mode showing more utterances than the telling one. No effects of the other factors and interactions were detected. The results of the model are reported in **Table 7**.

Next, we analyzed the mean length of children's utterances by calculating the number of words. As for the number of utterances, the model revealed a significant effect of only the Mode factor, with the retold stories containing a higher number of words than those produced in the Telling mode. The results of the model are reported in **Table 8**.

Taking the results together, the statistical analyses revealed a main effect of the Mode factor. The stories produced in the Retelling task showed higher scores in the Story Structure and the Comprehension questions and elicited a higher number of utterances with a higher mean length of utterance than the stories produced in the Telling task. In addition, Age had an effect on the Story Structure and the Comprehension questions but not on the Syntactic Measures: 5-year-old children obtained better scores than their 3- and 4-year-old peers. Quantity of Input did not yield any significant effect. Importantly, no effect was detected in the language of the experiment, whether Italian or Vicentino.

However, it should be stressed that, when looking at their production in the two versions of the tasks, we noticed that the 33 children completing both versions never produced an entire story in dialect. Despite understanding the comprehension

TABLE 4 | Descriptive statistics for Story Comprehension in bilingual MAIN-narratives across Age groups: mean and (standard deviation).

	Age	Italian		Vicentino	
		MAIN-Retelling	MAIN-Telling	MAIN-Retelling	MAIN-Telling
Comprehension	3	5.8 (1.2)	3.5 (0.9)	4.5 (1)	3.7 (0.8)
	4	6.1 (1.1)	3.9 (1)	4.6 (1.6)	3.7 (1.6)
	5	6.9 (1.8)	5.2 (1.8)	6.9 (1.6)	5.4 (1.5)

TABLE 5 | Results of the linear mixed model for the dependent variable Story Comprehension.

Predictors	df	*F*	*p*
Narrative-Mode	1, 26.61	39.282	< 0.001
Language-Mode	1, 18.52	1.042	0.320
Age	2, 18.03	7.283	0.005
Quantity of Input	6, 18.03	1.016	0.446
Narrative-Mode * Language-Mode	1, 36.00	5.330	0.067
Narrative-Mode * Age	2, 26.61	0.465	0.633
Language-Mode * Age	2, 18.52	1.022	0.379
Narrative-Mode * Quantity of Input	6, 26.61	0.946	0.479
Language-Mode * Quantity of Input	6, 18.52	0.748	0.618
Age * Quantity of Input	6, 18.03	0.818	0.570
Narrative-Mode * Language-Mode * Age	2, 36.00	0.533	0.591
Narrative-Mode * Language-Mode * Quantity of Input	6, 36.00	1.378	0.250
Narrative-Mode * Age * Quantity of Input	6, 26.61	1.057	0.412
Language-Mode * Age * Quantity of Input	6, 18.52	0.236	0.959
Narrative-Mode * Language-Mode * Age * Quantity of Input	6, 36.00	0.485	0.815

The model was fitted using restricted maximum likelihood. Full model summary: N = 132, REML 422.6.

TABLE 6 | Descriptive statistics for Syntactic Measures in bilingual MAIN-narratives across Age groups: mean and (standard deviation).

		Age	Italian		Vicentino	
			MAIN-Retelling	MAIN-Telling	MAIN-Retelling	MAIN-Telling
Syntax	N utterances	3	13.2 (6.2)	8.2 (5)	11 (6.4)	6.5 (1.6)
		4	13.3 (4)	8.6 (1.9)	12.5 (4.4)	10.9 (6.4)
		5	14.2 (2.9)	10.9 (3.9)	14.4 (7)	10.9 (4.5)
	MLU (words)	3	6.6 (0.9)	5.6 (0.9)	5.9 (0.8)	5.2 (1)
		4	6.9 (1.8)	5.3 (0.8)	6.2 (1)	5.8 (1)
		5	6.5 (0.9)	6.3 (1)	6.3 (0.8)	5.9 (0.6)

questions and the instructions in Vicentino, their replies and their narratives were mainly in Italian. Out of 2440 words produced in the Vicentino Retelling task, we counted 86 dialectal words, 1722 words in Italian, and 632 words which we could not determine as Italian or Vicentino. Out of 2037 words produced in the Vicentino Telling task, we counted 41 words that we could classify as dialectal and 1523 words in Italian. We could not assign a value to 473 words. In the Italian experiments, the presence of dialectal features was even lower. Out of 2607 words in the Italian Retelling task, 4 words were Vicentino, while 1832 were clearly Italian. In the Italian Telling task, 8 out of 2073 words were dialectal and 1788 were Italian.

This shows that there is a discrepancy between comprehension and production of the dialect. While all 33 children comprehended the dialect, not all of them produced Vicentino.

TABLE 7 | Results of the linear mixed model for the dependent variable Number of Utterances.

Predictors	df	*F*	*p*
Narrative-Mode	1, 18.03	12.439	0.002
Language-Mode	2, 18.11	1.375	0.278
Age	1, 18.57	0.861	0.365
Quantity of Input	6, 18.11	1.478	0.241
Narrative-Mode * Language-Mode	2, 18.03	0.317	0.733
Narrative-Mode * Age	2, 18.57	0.508	0.610
Language-Mode * Age	1, 36.00	0.108	0.744
Narrative-Mode * Quantity of Input	6, 18.11	2.042	0.112
Language-Mode * Quantity of Input	6, 18.03	0.352	0.900
Age * Quantity of Input	6, 18.57	0.658	0.684
Narrative-Mode * Language-Mode * Age	2, 36.00	1.086	0.348
Narrative-Mode * Language-Mode * Quantity of Input	6, 18.03	0.677	0.670
Narrative-Mode * Age * Quantity of Input	6, 18.57	1.381	0.273
Language-Mode * Age * Quantity of Input	6, 36.00	2.230	0.062
Narrative-Mode * Language-Mode * Age * Quantity of Input	6, 36.00	1.257	0.302

The model was fitted using restricted maximum likelihood. Full model summary: N = 132, REML 568.8.

TABLE 8 | Results of the linear mixed model for the dependent variable MLU.

Predictors	df	*F*	*p*
Narrative-Mode	1, 21.58	32.726	< 0.001
Language-Mode	1, 18.02	0.858	0.367
Age	2, 18.04	1.451	0.260
Quantity of Input	6, 18.04	0.763	0.608
Narrative-Mode * Language-Mode	1, 36.00	0.069	0.795
Narrative-Mode * Age	2, 21.58	1.526	0.240
Language-Mode * Age	2, 18.02	0.312	0.736
Narrative-Mode * Quantity of Input	6, 21.58	2.323	0.070
Language-Mode * Quantity of Input	6, 18.02	2.067	0.109
Age * Quantity of Input	6, 18.04	0.373	0.887
Narrative-Mode * Language-Mode * Age	2, 36.00	4.549	0.067
Narrative-Mode * Language-Mode * Quantity of Input	6, 36.00	2.150	0.071
Narrative-Mode * Age * Quantity of Input	6, 21.58	1.668	0.177
Language-Mode * Age * Quantity of Input	6, 18.02	0.440	0.842
Narrative-Mode * Language-Mode * Age * Quantity of Input	6, 36.00	0.659	0.683

The model was fitted using restricted maximum likelihood. Full model summary: N = 132, REML 355.

Linguistic Features

Next, we analyzed children's productions from a qualitative viewpoint to detect the amount and type of dialectal elements present in the stories. On the basis of the dialectal elements produced by the 33 children, we delineated three profiles.

The first profile comprises five children who comprehended the instructions and the questions in Vicentino but did not produce any dialectal element at the phonological, lexical, morphological, or syntactic level. Interestingly, in the parental questionnaires, these 5 children were reported to exclusively use Italian in their interactions with the caregivers. In addition, the quantity of dialectal input these children were

reported to receive was either null or very scarce (see **Appendix Tables A,B**).

The remaining 28 children produced a story mainly in Italian but with some dialectal interference present. All 28 children produced some dialectal elements in the Vicentino experiment and did so more in the Retelling story (N = 106) than in the Telling story (N = 41). Three children also produced 12 dialectal elements in the Italian Telling and Retelling tasks: these were mainly lexical elements like *toso*, *cana*, *bala*, but in four occurrences children produced the dialectal auxiliary selection, as in *se l'ha ripreso* 'he took it back.'

Among the 28 children who produced some Vicentino elements, 19 children comprehended the instructions and the questions and produced phonological, morphological, and lexical dialectal elements. This represents the second profile. All of them were reported to exclusively or mainly use Italian in their interactions with the caregivers (see **Appendix Table A**). Eleven children were exposed to some dialect for at least 12% of the interaction time with the caregiver, while the remaining eight children were reported to receive very little or null dialectal input (see **Appendix Table B**).

Among the phonological dialectal features, all these children produced apocope of the final vowels/e/and/o/following/n/as illustrated in (4), with the consequent velarization of the nasal consonant. Out of 560 produced words that contained potential contexts for apocope to apply, 84 words were produced with apocope. In 38 cases we were able to determine that children also produced the final nasal consonant as a velar nasal.

(4) (a) [cà:ŋ]
dog
(b) [balò:ŋ]
balloon
(c) [mà:ŋ]
hand
(d) [bambì:ŋ]
young boy

As noticed in the dialectological literature, in Vicentino/e/and/o/occurring in word-final position do not undergo apocope when they are preceded by other consonants, and neither do/r/nor/l/, unlike the Venetian dialect. Interestingly, all children respected this rule as shown by the productions in (5) where /e,o/ follow consonants different from/n/and are not dropped.

(5) (a) [albà:ro]
tree
(b) [farfà:le]
butterflies

In intervocalic context/t, l/ are degeminated (6a,b) and voiceless plosives become voiced as in (6c,d). For those items where degemination could have occurred, children always produced it. Children never produced the dialectal words without degemination, e.g., *balla* instead of *bala*. Children either produced it with degemination or they produced the Italian lexeme *palla*, without degemination.

(6) (a) *bala*
'ball' (It. *palla*)
(b) *gato*
'cat' (It. *gatto*)
(c) *bugo*
hole (It. *buco*)
(d) *morsegato*
bitten (It. *morsicato*)

As for the morphological dialectal features, all children produced the Vicentino weak form of the masculine singular article/el/: *el can* 'the dog,' *el balon* 'the balloon,' *el toso* 'the young boy.' Out of a total of 422 definite articles, 286 were Vicentino, while the remaining articles were Italian. Sometimes, the Vicentino article also introduced Italian nouns (N = 45), as in *el cane* 'the dog.'

Some children also produced predicates with clear dialectal prefixes. One case is represented by the verb *incorzarse* 'to realize' with the prefix *in-* while the corresponding Italian predicate has the prefix *a-*, *accorgersi*: *el can se incorze che* [. . .] 'the dog realizes that [. . .]'. Another case is the verb *scumissiare* 'to begin, start' with the prefix *s-* instead of the corresponding Italian *in-*, *incominciare*.

The dialect was also present in other lexical items. For instance, all 19 children used the verb *ciapare* 'to take, catch' at least once: *el can voe ciapare el topo* 'the dog wants to catch the mouse.' In a few cases, some children also used the verb in the idiomatic use with the noun *paura* 'scare,' as in *el toso ciapa paura* 'the boy got scared.'

Finally, the third profile comprises nine children who comprehended the instructions and the questions in dialect and, in addition to phonological and lexical dialectal elements, also produced syntactic dialectal structures. Overall, the profile matched the expectations we had, based on the parental questionnaire. This especially holds for three children who were reported to mainly receive dialectal input and in turn to reply in Vicentino. Likewise, two children were reported to receive input and reply in Vicentino for 33% of their interaction time with the caregivers. However, for four children our analysis diverged from the results of the parental questionnaire: they were reported to be exposed to dialect for at least 33% of their child–caregiver interaction time but to use Italian either exclusively or almost exclusively (see **Appendix Table B**). In addition to the phenomena quoted for the second profile, these nine children produced dialectal syntactic structures, all of which occurred in code-switched utterances. Interestingly, the dialectal syntactic structures produced by the nine children share one commonality: they are all related to the lower portion of the clause, involving aspectual, voice, and tense functional projections. Conversely, subject clitics and the locative/dative clitic *ghe*, which are clear syntactic properties of Vicentino, were not present in any of the children's productions. Likewise, complementizers were all

produced in Italian: *quando* 'when' instead of the dialect *cò, perché* 'because' instead of *parché, come* 'as' instead of *come che*.

In at least one instance, all children produced particle verbs. Some examples are *tornare zò* 'lit. to go back down; to climb down,' *saltare drio* 'lit. to jump behind; to chase,' *corere drio* 'lit. to run behind; to chase,' *abaiare drio* 'lit. to bark behind; to bark,' *molare zò* 'lit. to release down; to release.' The particle verb was also produced when the phonological shape of the lexemes was Italian, as in the case of *va su per l'albero* 'he climbs the tree.'

Another domain in which dialectal production was detected was the selection of the auxiliary. With reflexive, impersonal, and modal verbs, Vicentino selects the auxiliary 'to be,' while Italian has the auxiliary 'to have.' Children produced the structures in (7) with the auxiliary verb 'to be.'

(7) (a) *se* *gavea* *fato* *male* *a-la* *testa*
cl.refl have.pst.3sg made bad a-the head
'He hurt his head.' (AS_04_T)

(b) *se* *lo* *gà* *riciapà*
cl.refl cl.3sg.m have.3sg taken_back
'He took it back.' (LG_04_SC)

(c) *se* *gà* *mangiato* *le* *luganeghe*
cl.refl have.3sg eaten the sausages
'He ate the sausages.' (MDS_59_T)

Finally, some children also produced the imperfective progressive construction typical of the Vicentino dialect and ungrammatical in Italian. The periphrasis is formed by the auxiliary *essere* 'to be,' the adverb *drio* 'behind' and a bare infinitive. This is illustrated in (8). Notice that in (8a), the structure is dialectal as well as the adverb *drio*, but the complementizer, the null subject, and the lexical predicate are Italian. Likewise, in (8b), the periphrasis is dialectal: the adverb and the lexical predicate are Vicentino, but the null subject and the form of the direct object matches with Italian.

(8) (a) *quando* *è* *drio* *prender-lo*
when be.3sg behind take-cl.3sg.m
'When he is grasping it' (AS_04_T)

(b) *È* *drio* *ciapare* *un* *topo*
be.3sg behind take a mouse
'He is chasing a mouse.' (SM_04_T)

DISCUSSION

This paper investigated children's linguistic competence in a standard language, namely Italian, and a vernacular non-standardized local dialect, namely the Vicentino dialect. Forty-four children from age 3–5 were tested with two narrative tasks in both Italian and Vicentino. We asked four questions: (Q1) Does children's performance in narrative comprehension and production differ depending on the mode of the narrative task, i.e., retelling vs. telling?; (Q2) Are there differences in narrative abilities between children determined by age?; (Q3) Does children's performance in narrative comprehension and production differ depending on the language of the experiment, i.e., Vicentino vs. Italian?; (Q4) Does children's competence depend on the quantity of input?.

(Q1) The Role of Narrative Mode

We found a significant effect of the narrative mode on both the macrostructure and the microstructure measures, showing that retelling elicited more coherent, articulated and longer stories than telling. In addition, the retelling task enhanced the comprehension of the story. Our results met our expectations and are in line with previous findings on bilinguals. Various studies on bilinguals found that bilingual children comprehended better and produced more structured and coherent stories with the Retelling than with the Telling task (e.g., Kunnari et al., 2016; Maviş et al., 2016; Roch et al., 2016; Bonifacci et al., 2018; Levorato and Roch, 2020; Roch and Hržica, 2021). While the literature generally agrees that the Retelling task enhances the macrostructure, the results on the role of the narrative mode for the microstructure are more controversial. Some studies found that presenting the child with a model story improved the lexical and syntactic complexity of the story (e.g., Adlof et al., 2014). Conversely, others reported that the microstructure improved in the Telling task (e.g., Gutiérrez-Clellen, 2002). In our case, the retold stories contained more utterances and exhibited a longer MLU. Hence, Retelling seems to positively affect these two syntactic measures. This observation also holds for the number of dialectal elements elicited, which was higher in the Retelling task than in the Telling.

(Q2) The Role of Age

We found a main effect of age in the macrostructure, showing that the complexity and accuracy of children's narratives grow with children's ages. Three- and four-year-old children differed from their older peers in both story structure and comprehension. Our findings are in line with previous studies which demonstrated that children's narrative abilities increase with age (e.g., Pearson, 2002; Florit et al., 2014; Bohnacker, 2016; Bonifacci et al., 2018).

(Q3–4) The Role of Language-Mode and Input

The language of the experiment, Italian vs. Vicentino, did not yield any significant effect in children's responses in the macrostructure and, to some extent, in the microstructure measures as well. This result corroborates earlier findings on bilinguals, where macrostructure measures also remained relatively invariant across bilingual children's languages (Pearson, 2002; Fiestas and Peña, 2004; Gagarina et al., 2015; Kunnari et al., 2016). Story macrostructure has usually been claimed to be less dependent on language abilities as compared to microstructure (Berman and Slobin, 1994; Rodina, 2016). We also found no effect of language in the two microstructure measures, i.e., the number of utterances and the MLU. Notice, however, that when we looked into children's production, the majority of lexemes and structures in both experiments were Italian. As clarified in Section "Results," children responded mostly in Italian in the Venetan

mode. The overabundance of Italian in the Vicentino experiment may be the reason why we did not find substantial differences in the microstructure measures across the two experiments. If the children had responded in Venetan, there might have been differences. Yet, it is interesting to note that, even if children's productions mainly consisted of an Italian lexicon and morphosyntax, the use of Vicentino in the instruction, the model story, and the comprehension questions did not negatively affect children's production. A clear difference due to the language-mode emerged in the linguistic feature analysis, to which we will return later.

As seen for language, quantity of input was also not detected as a significant factor in the macrostructure. Our result is in line with previous findings by Gagarina (2016) on Russian–German bilinguals, Bohnacker (2016) on Swedish–English bilinguals, Kunnari et al. (2016) on Finnish–Swedish children, Iluz-Cohen and Walters (2012) on English–Hebrew, and Rodina (2016) on Russian–Norwegian bilinguals. These studies found that narrative macrostructure is relatively invariant across languages and is not much reliant on language proficiency, while narrative microstructure seems to be more dependent on language proficiency. However, we also find no effect of language exposure in the microstructure measures. As for the language-mode factor, this finding may be explained by children's overuse of Italian in the Vicentino experiment. Indeed, children's productions mainly consisted of Italian lexical items and syntactic structures in both experiments. Conversely, the results on the parental questionnaire showed that the quantity of dialectal input children are exposed to positively correlates with the quantity of dialect children were reported to use. Likewise, the linguistic profiles drawn from the parental questionnaires generally matched the results on the linguistic feature analysis, suggesting that language exposure does play a role in the production of dialect, as suggested in previous studies on Venetan-Italian children (Klaschik and Kupisch, 2016; Kupisch and Klaschik, 2017).

Our study further showed that all the 44 preschool children that were reported to have been exposed to dialect and Italian from birth produced and comprehended Italian. Out of them, 11 children were classified as monolingual Italian speakers since they failed to understand the questions in the familiarization phase or the experimental instructions in Vicentino. Interestingly, in the parental questionnaire, these children were reported to use Italian exclusively or almost exclusively in their interactions with the caregivers. Among them, 10/11 children were reported to receive scarce or null input in the Vicentino dialect in the interactions with their caregivers. We found one exception: one child did not understand the dialect questions although she was reported to receive dialectal input 43% of the time. Overall, we can conclude that quantity of input plays a role in the comprehension of the dialect.

The remaining 33 children comprehended Vicentino. Depending on the quantity and type of Vicentino elements in their productions, we identified three linguistic profiles. The first profile is represented by receptive bilinguals (Mioni, 1979), comprising five children who comprehended the instructions and the questions in Vicentino but did not produce any dialectal element at the phonological, lexical, morphological, or syntactic level. Interestingly, in the parental questionnaire, these five children were reported to exclusively use Italian in their interactions with the caregivers and to receive either null or very scarce quantity of dialectal input. For these children as well, the quantity of exposure seems to play a role in the production of dialect. The second profile is represented by 19 bilinguals who mainly produced their stories and replied in Italian but with phonological and morphological dialectal elements. For this group of children, we did not find any stable link with the quantity of input reported in the parental questionnaire: all children were reported to exclusively or mainly use Italian in their interactions with the caregivers; 11 children were exposed to some dialect for at least 12% of the interaction time with the caregiver; 8 children were reported to receive very little or null dialectal input. The third profile is represented by nine children who produced some dialectal elements at the phonological and morphological level, as in the second profile, but also at the syntactic level. Overall, the children in this profile matched the expectations we had, based on the results from the parental questionnaire. All children were reported to mainly receive dialectal input or in at least 33% of their interaction time with the caregivers (see **Appendix Table B**).

We may conclude that, with some exceptions, the quantity of Vicentino input children receive has a positive effect on children's comprehension and production abilities of the minority language. Our results nicely match those on Venetan in Kupisch and Klaschik (2017). Infrequent dialect users were found to rely more on Italian than frequent dialect users. The authors suggested that children's dialectal production depends on the quantity and quality of the dialectal input. Our data nicely confirm this conclusion. Conversely, our results diverge from the findings on Friulian in Covazzi (2019): the author did not find a correlation between children's Friulian production and the quantity of Friulian input reported in the parental questionnaires. Various reasons, both methodological and linguistic, may account for this difference. One factor may be the different sizes of the participants' samples: while we tested 44 children, Covazzi analyzed the production of 23 children aged 4-to-6 years. Given the high degree of variation we and Covazzi as well found in the parental responses, a smaller sample may not be sufficient to draw any generalization. Another reason may lie in the scale used to evaluate the quantity of input: while our scale had 10 points, Covazzi adopted a scale with four levels, i.e., 'almost always,' 'often,' 'sometimes,' 'almost never.' Participants may have faced difficulties in quantifying the dialectal input with a non-numerical scale. Finally, the differences between our and Covazzi's results may reflect the different sociolinguistic profiles of Friulian and Venetan (Vanelli, 2005; Vicario, 2015).

The linguistic analysis also suggests an implicational scale in the dialectal competence: if a child exhibits some productions with dialectal syntax, s/he also produces dialect at the phonological, morphological, and lexical level, but not vice versa. From a theoretical perspective, it is interesting to note that the dialectal syntactic structures children produced are all related to the lower portion of the clause, involving verbal, voice,

and aspectual functional projections. On the other hand, the higher layers of the clause, where subject and locative clitics, as well as complementizers, are merged, are only Italian. This result is consistent with the Growing Tree Hypothesis proposed by Friedmann et al. (2020). According to this view, children's developmental stages follow the geometry of the syntactic tree (see Rizzi, 2004): early stages correspond to small portions of the adult syntactic tree, which grows during development. Although further studies are in order, it seems plausible to extend this approach to bilectal acquisition as well.

CONCLUSION

Our results demonstrated that dialectal competence is already present in preschool children. In addition, our findings suggest that dialectal competence should be arranged along a fine-grained continuum. Unlike "standard" bilingual speakers, bilectal speakers can comprehend dialect although they may completely lack the competence to produce it. In this sense, they can be qualified as receptive bilinguals. Bilectal speakers may also produce some dialectal elements. Some bilectal children only have access to the phonological, morphological, and lexical domain of the dialect, while others also produce dialectal syntactic structures. Although the results on the macrostructure are similar to the findings from standard bilingual studies, the strikingly small number of dialectal elements produced by children suggests that bilectal acquisition may be different from standard bilingualism. This may be a reflection of the sociolinguistic differences outlined in Section "Introduction," but also of the linguistic challenges discussed in Section "Analysis and scoring." As reported in previous studies on bilectal acquisition, there is indeed a high degree of linguistic overlap between Vicentino and Italian which regards the lexical, phonological, morphological, and syntactic levels. As a result, it is not straightforward to determine whether a production qualifies as dialectal or Italian and in turn to provide an appropriate measure of how much dialect children produce. For future work, it would be relevant to test bilectal children with different experimental methods, such as grammaticality judgment tasks, to tease apart the lack of production of a given structure from the lack of grammatical competence of that structure. Moreover, since the dialectological profiles in Italy differ from one region to another, future studies on different dialects are necessary to establish the role of the extra-linguistic factors on bilectal acquisition. As a matter of fact, our results may be extended to other regions with a context similar to Veneto, maybe Apulia and Basilicata, but not necessarily to others, like Liguria or Lombardy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

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

REFERENCES

Adlof, S. M., McLeod, A. N., and Leftwich, B. (2014). Structured narrative retell instruction for young children from low socioeconomic backgrounds: a preliminary study of feasibility. *Front. Psychol.* 5:391. doi: 10.3389/fpsyg.2014.00391

Antoniou, K., Grohmann, K. K., Kambaranos, M., and Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition* 149, 18–30. doi: 10.1016/j.cognition.2015.12.002

Benincà, P. (1994). *La variazione Sintattica: Studi di Dialettologia Romanza*. Bologna: Il Mulino.

Benincà, P., and Poletto, C. (2006). "Phrasal verbs in venetan and regional Italian," in *Language Variation – European Perspectives*, ed. F. Hinskens (Amsterdam: John Benjamins), 9–22. doi: 10.1075/silv.1.02ben

Benincà, P., and Vanelli, L. (1982). "Appunti di sintassi veneta," in *Guida ai Dialetti Veneti IV*, ed. M. Cortelazzo (Padova: CLEUP), 7–38.

Berman, R. A., and Slobin, D. I. (1994). "Narrative structure," in *Relating Events in Narrative: A Crosslinguistic Developmental Study*, eds R. A. Berman and D. I. Slobin (Hillsdale, NJ: Lawrence Erlbaum), 39–54.

Berruto, G. (2005). "Dialect/standard convergence, mixing, and models of language contact: the case of Italy," in *Dialect Change. Convergence and Divergence in European Languages*, eds P. Auer, F. Hinskens, and P. Kerswill (Cambridge: Cambridge University Press), 81–97. doi: 10.1017/CBO9780511486623.005

Berruto, G. (2006). "Quale dialetto per l'Italia del duemila? Aspetti dell'italianizzazione e 'risorgenze' dialettali in piemonte (e altrove)," in *Lingua e Dialetto nell'Italia del Duemila*, eds A. Sorbero and A. Miglietta (Galatina: Congedo Editore), 117–123.

Berruto, G. (2007). *Lingue Minoritarie e Sociolinguistica Del Contatto*. Roma: Carocci.

Bohnacker, U. (2016). Tell me a story in english or swedish: narrative production and comprehension in bilingual preschoolers and first graders. *Appl. Psychol.* 37, 19–48. doi: 10.1017/S0142716415000405

Bonifacci, P., Barbieri, M., Tomassini, M., and Roch, M. (2018). In few words: linguistic gap but adequate narrative structure in preschool bilingual children. *J. Child Lang.* 45, 120–147. doi: 10.1017/S0305000917000149

Bonifacio, E. (2014). *The Acquisition of the Copula in two Bilingual Italian-Venetan Children*. MA thesis. Venice: University Ca' Foscari.

Cardinaletti, A. (2013). On the bilingual acquisition of Italian and venetan dialects: a focus on subject and object clitic pronouns. *Paper Presented at the Generative Linguistics in the Old World (Glow) 36, April 2-6 2013, Lund University*, Lund.

Cerruti, M. (2011). Regional varieties of Italian in the linguistic repertoire. *Int. J. Sociol. Lang.* 210, 9–28. doi: 10.1515/ijsl.2011.028

Cortelazzo, M. (1981). *Guida ai Dialetti Veneti*. Padova: CLEUP.

Covazzi, C. (2019). *Italian-Friulian Children: An Investigation into their Bilingualism*. Doctoral thesis. Padova: University of Padua.

Davies, A. (2003). *The Native Speaker: Myth and Reality*, 2nd Edn. Clevedon: Multilingual Matters. doi: 10.21832/9781853596247

Davies, A. (2013). *Native Speakers and Native Users: Loss and Gain*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139022316

Dicataldo, R., Florit, E., and Roch, M. (2020). Fostering broad oral language skills in preschoolers from low SES background. *Int. J. Environ. Res. Public Health* 17:4495. doi: 10.3390/ijerph17124495

Dicataldo, R., and Roch, M. (2020). Are the effects of variation in quantity of daily bilingual exposure and socioeconomic status on language and cognitive abilities

independent in preschool children? *Int. J. Environ. Res. Public Health* 17:4570. doi: 10.3390/ijerph17124570

Fiestas, C. E., and Peña, E. D. (2004). Narrative discourse in bilingual children: language and task effects. *Lang. Speech Hear. Serv. Sch.* 35:155. doi: 10.1044/0161-1461(2004/016)

Florit, E., Roch, M., and Levorato, M. C. (2014). Listening text comprehension in preschoolers: a longitudinal study on the role of semantic components. *Read. Writ. Interdiscip. J.* 27, 793–817. doi: 10.1007/s11145-013-9464-1

Friedmann, N., Belletti, A., and Rizzi, L. (2020). *Growing Trees: The Acquisition of the Left Periphery*. Siena: Ms. University of Tel Aviv.

Gagarina, N. (2016). Narratives of Russian–German preschool and primary school bilinguals: rasskaz and erzaehlung. *Appl. Psychol.* 37, 91–122. doi: 10.1017/S0142716415000430

Gagarina, N., and Bohnacker, U. (2022). *Storytelling in Bilingual Children: How Organization of Narratives is (not) Affected by Linguistic Skills and Environmental Factors. Linguistic Approaches to Bilingualism.*

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., et al. (2012). *MAIN: Multilingual Assessment Instrument for Narratives. ZAS Papers in Linguistics*. Berlin: ZAS, 56. doi: 10.21248/zaspil.56.2019.414

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, 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

Gagarina, N., and Lindgren, J. (2020). New language versions of MAIN: multilingual assessment instrument for narratives – revised. *ZAS Papers Ling.* 64:274. doi: 10.21248/zaspil.64.2020.543

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

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: assessing performance of bilingual children. *Ling. Educ.* 13, 175–197. doi: 10.1016/S0898-5898(01)00061-4

Hoff, E. (2015). "Language development in bilingual children," in *The Cambridge Handbook of Child Language*, 2nd Edn, eds E. Bavin and L. Naigles (Cambridge: Cambridge University Press), 483–503. doi: 10.1017/CBO9781316095829.022

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., and Welsh, S. N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: a longitudinal study from two to four years. *Early Child. Res. Q.* 29, 433–444. doi: 10.1016/j.ecresq.2014.04.012

Houston, D. M., and Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *J. Exp. Psychol. Hum. Percep. Perform.* 26, 1570–1582. doi: 10.1037/0096-1523.26.5.1570

Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-Native Speakers: Theory and Research*. Amsterdam: John Benjamins. doi: 10.1075/lllt.41

Iluz-Cohen, P., and Walters, J. (2012). Telling stories in two languages: narratives of bilingual preschool children with typical and impaired language. *Bilingual. Lang. Cogn.* 15, 58–74. doi: 10.1017/S1366728911000538

ISTAT (2015). *La Lingua Italiana, i Dialetti e le Lingue Straniere*. Roma: Istituto nazionale di statistica.

Klaschik, E., and Kupisch, T. (2016). Overt and null subjects in the two varieties in Venetian-Italian school children. *Lingue Linguaggio* 2, 169–198.

Kunnari, S., Vañlimaa, T., and Laukkanen-Nevala, P. (2016). Macrostructure in the narratives of monolingual finnish and bilingual finnish–swedish children. *Appl. Psychol.* 37, 123–144. doi: 10.1017/S0142716415000442

Kupisch, T., and Klaschik, E. (2017). "Cross-lectal influence and gender marking in bilectal Venetan-Italian acquisition," in *Cross-Linguistic Influence in Bilingualism*, eds E. Blom, L. Cornips, and J. Schaeffer (Amersterdam: John Benjamins), 127–152. doi: 10.1075/sibil.52.07kup

Lauchlan, F., Parisi, M., and Fadda, R. (2012). Bilingualism in Sardinia and Scotland: exploring the cognitive benefits of speaking a 'minority' language. *Int. J. Bilingual.* 17, 43–56. doi: 10.1177/1367006911429622

Leivada, E., Papadopoulou, E., Kambaranos, M., and Grohmann, K. (2017). The influence of bilectalism and non-standardization on the perception of native grammatical variants. *Front. Psychol.* 8:205. doi: 10.3389/fpsyg.2017.00205

Levorato, M. C., and Roch, M. (2020). Italian adaptation of the multilingual assessment instrument for narratives. *ZAS Pap. Linguist.* 64, 139–146. doi: 10.21248/zaspil.64.2020.566

Loporcaro, M. (2013). *Profilo Linguistico dei Dialetti Italiani*. Bari: Editori Laterza.

Marcato, G. (2005). "Il rapporto tra "lingua" e "dialetto" come "spazio dinamico di variazione," in *Lingue e Dialetti nel Veneto*, Vol. 3, ed. G. Marcato (Padova: CLEUP), 18–20.

Marcato, G., and Ursini, F. (1998). *Dialetti Veneti. Grammatica e Storia*. Padova: CLEUP.

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

Mioni, A. (1979). "La situazione sociolinguistica italiana: lingua, dialetto, italiani regionali," in *Guida All'educazione Linguistica*, ed. A. Colombo (Bologna: Zanichelli), 101–114.

Mioni, A. (2007). "Considerazioni generali, a partire dalla situazione veneta," in *La Dialectologie aujourd'hui. Atti del Convegno Internazionale «Dove va la dialettologia?»*, eds G. Raimondi and L. Revelli (Alessndria: Edizioni dell'Orso), 149–154.

Mioni, A., and Trumper, J. (1977). "Per un'analisi del continuum linguistico veneto," in *Aspetti Sociolinguistici dell'Italia Contemporanea*, eds R. Simone and G. Ruggiero (Roma: Bulzoni), 329–371.

Pearson, B. Z. (2002). "Narrative competence among monolingual and bilingual school children in miami," in *Language and Literacy Development in Bilingual Children*, eds D. Kimbrough Oller and R. E. Eilers (Clevedon: Multilingual Matters), 135–174. doi: 10.21832/9781853595721-008

Pellegrini, G. B. (1977). *Carta dei Dialetti d'Italia*. Pisa: Pacini.

Poletto, C. (1996). "Three types of subject clitics and the theory of pro," in *Parameters and Functional Heads*, eds A. Belletti and L. Rizzi (Oxford: Oxford University Press), 269–300.

Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris. doi: 10.1515/9783110883718

Rizzi, L. (2004). "Locality and left periphery," in *Structures and Beyond – the Cartography of Syntactic Structures*, Vol. 3, ed. A. Belletti (Oxford: Oxford University Press), 223–251.

Roch, M., Florit, E., and Levorato, M. C. (2016). Narrative competence of italian-english bilingual children between 5 and 7 Years. *Appl. Psychol.* 37, 49–67. doi: 10.1017/S0142716415000417

Roch, M., and Hržica, G. (2021). "Narrative comprehension by croatian. Italian Bilingual children aged 5-7 years old. The role of receptive vocabulary and sentence comprehension," in *Developing Narrative Comprehension. Multilingual Assessment Instrument for Narratives*, eds U. Bohnacker and N. Gagarina (Amsterdam: John Benjamins Publishing Company), 172–196. doi: 10.1075/sibil.61.06roc

Rodina, Y. (2016). Narrative abilities of preschool Bilingual Norwegian-Russian children. *Int. J. Bilingualism* 21, 617–635. doi: 10.1177/1367006916643528

Santipolo, M., and Tucciarone, S. (2006). ""Dalla semidialettofonia di ritorno al bilinguismo consapevole: un'ipotesi di evoluzione sociolinguistica in Veneto," in *Lingue in Contatto a Scuola. Tra italiano, Dialetto e Italiano L2*, eds I. Tempesta and M. Maggio (Milano: FrancoAngeli), 162–168.

Sorace, A. (2003). "Near-nativeness," in *The Handbook of Second Language Acquisition*, eds C. J. Doughty and M. H. Long (Oxford: Blackwell), 130–151. doi: 10.1002/9780470756492.ch6

Tessarolo, M., and Bordon, E. (2015). "Perché gli adulti si rivolgono ai bambini in italiano?," in *Dialetto. Parlato, Scritto, Trasmesso*, ed. G. Marcato (Padova: CLEUP), 59–65.

Vanelli, L. (2005). "La posizione del friulano nel dominio romanzo," in *Linguistica Friulana*, eds P. Benincà and L. Vanelli (Padova: Unipress), 19–31.

Vangsnes, ØA., Söderlund, G. B., and Blekesaune, M. (2015). The effect of bidialectal literacy on school achievement. *Int. J. Bilingual Educ. Bilingual.* 20, 346–361. doi: 10.1080/13670050.2015.1051507

Vicario, F. (2015). "Friulian dialects," in *Handbook of Friulian linguistics*, eds F. Fabbro and C. Crescentini (Udine: Forum), 161–172.

Zamboni, A. (1974). *Veneto. Profilo dei Dialetti Italiani 5*. Pisa: Pacini.

Zamboni, A. (1979). "Le caratteristiche essenziali dei dialetti veneti," in *Guida ai Dialetti Veneti*, ed. M. Cortelazzo (Padova: CLEUP), 9–44.

Zamboni, A. (1989). "Pavano e padovano tra continuità e innovazione," in *Secondo Convegno Internazionale di Studi sul Ruzante*, eds G. Calendola and G. Vellucci (Venezia: Corbo e Fiore), 273–283.

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

APPENDIX

TABLE A | Overview of the linguistic profiles of the children according to children's input and output in Italian reported in the Italian version of The Questionnaire for Parents of Bilingual Children (Italian version: Dicataldo and Roch, 2020; Levorato and Roch, 2020).

Italian input	Italianoutput	N of children
10	10	13
9.6	10	2
9.5	9.5	1
9.5	10	1
9.2	10	2
9	10	1
8.8	10	3
8.6	9.2	1
8.5	10	1
8.4	10	2
8.3	10	1
8	10	3
7.2	10	1
7.2	8.5	1
7	10	1
6.8	9	1
6.7	6.7	1
6.7	10	1
6.6	8.4	1
6.5	9	1
5.7	9	1
4.3	1	1
3	1.5	1
2.3	2	1
1.5	10	1

TABLE B | Overview of the linguistic profiles of the children according to children's input and output in dialect reported in the Italian version of The Questionnaire for Parents of Bilingual Children (Italian version: Dicataldo and Roch, 2020; Levorato and Roch, 2020).

Dialectal input	Dialectaloutput	N of children
0	0	13
0.4	0	2
0.5	0.5	1
0.5	0	1
0.8	0	2
1	0	1
1.2	0	3
1.4	0.8	1
1.5	0	1
1.6	0	2
1.7	0	1
2	0	3
2.8	0	1
2.8	1.5	1
3	0	1
3.2	1	1
3.3	3.3	1
3.3	0	1
3.4	1.6	1
3.5	1	1
4.3	1	1
5.7	9	1
7	8.5	1
7.7	8	1
8.5	9	1

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fpsyg.2021.717635

The 'Comparative Logic' and Why We Need to Explain Interlanguage Grammars

Laura Domínguez[1]***and María J. Arche*[2]

[1]Modern Languages and Linguistics, University of Southampton, Southampton, United Kingdom, [2]School of Humanities and Social Sciences, Centre for Research and Enterprise in Language, University of Greenwich, London, United Kingdom

In this paper we argue that Bley-Vroman's Comparative Fallacy, which warns against comparisons between native speakers and learners in second-language acquisition (SLA) research, is not justified on either theoretical or methodological grounds and should be abandoned as it contravenes the explanatory nature of SLA research. We argue that for SLA to be able to provide meaningful explanations, grammatical comparisons with a baseline (usually of native speakers although not always the case) are not only justified but necessary, a position which we call the 'Comparative Logic'. The methodological choices assumed by this position ensure that interlanguage grammars are analysed in their own right and respecting their own principles. Related issues, such as why we focus on the native speaker and why investigating deficits in linguistic-cognitive SLA is essential in our field are discussed as well.

Keywords: comparative fallacy, native speaker, interlanguage, control group, Universal Grammar

OPEN ACCESS

Edited by:
Antonella Sorace, University of Edinburgh, United Kingdom

Reviewed by:
Jacee Cho, University of Wisconsin-Madison, United States
Alejandro Javier Wainselboim, CONICET Mendoza, Argentina

****Correspondence:***
Laura Domínguez
ldo@soton.ac.uk

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

Received: *31 May 2021*
Accepted: *13 September 2021*
Published: *25 October 2021*

Citation:
Domínguez L and Arche MJ (2021) The 'Comparative Logic' and Why We Need to Explain Interlanguage Grammars. Front. Psychol. 12:717635. doi: 10.3389/fpsyg.2021.717635

THE NATURE OF COMPARISONS IN SECOND-LANGUAGE ACQUISITION RESEARCH

The view that comparisons between native speakers (NS) and non-native speakers (NNS), which are pervasive in second-language acquisition (SLA) research, should be discouraged is not new (Firth and Wagner, 1997; Klein, 1998). Recently, however, concerns about the use of these comparisons have been raised among some researchers working within the so-called linguistic-cognitive approaches to SLA[1] arguing that comparing learners with natives falls into a 'comparative fallacy' (*CF*) as described by Bley-Vroman (1983) and help promote a monolingual bias in our field. The *CF* rests on two key claims: (1) the linguistic system of the learner [the interlanguage grammar (ILG)] is a system on its own right and (2) comparisons between ILG and other systems (including the target grammar) are not legitimate under any circumstances (see also Lakshmanan and Selinker, 2001). We argue, however, that these two claims are independent from each other. Indeed, many SLA researchers have explicitly claimed that the language of the second-language acquirers 'represents a linguistic system in its own right and should be investigated as such (Huebner, 1983, p. 33)'; this view is consistent with Selinker's (1972)

[1]Linguistic-cognitive approaches include researchers working on cognitive-interactionist, instructed, psycholinguistically oriented, usage-based and Generative SLA according to Ortega (2014, 2019) who claims that their '*main disciplinary goal is to illuminate the human capacity for language, and most would also share post-positivist logics, quantitative rigor, and generalizability as values in their research* (Ortega, 2019, p. 23)'.

original conception of 'interlanguage', and it is widely accepted in our field. The main concern for researchers (from all fields) is the legitimacy of NS-NNS comparisons (Firth and Wagner, 1997; Klein, 1998; Cook, 1999).[2]

According to the *CF*, ILG should be investigated without being compared with NS (the control group) as this may result in a view of learner grammars as 'degenerate' systems of less quality (i.e., the comparison necessarily presupposes a prejudice against NNS). In this paper, we position ourselves against this view (see also Mack, 1997; Montrul, 2013) and argue that, despite its increasing popularity among linguistic-cognitive SLA researchers, the *CF* is not justified on either theoretical or methodological grounds. Furthermore, we propose replacing the *CF* with the Comparative Logic which justifies comparisons with a baseline as these not only allow researchers to study L2 grammars 'in their own right' but are also essential in order to examine and explain the nature of L2 grammars.

In this paper, we will present and defend the Comparative Logic on the basis of the following arguments:

The Comparative Fallacy Is Routinely Misunderstood

The methodological decisions to prevent the *CF* entail much more than not including a control group of native speakers in the design of a study. We will argue that the *CF*, in fact, constitutes a significant barrier to providing meaningful analyses and explanations, and it does not support the fundamental explanatory nature of the field.

Acquiring a Language Is an Incremental Process and Learners' Grammars Develop Towards a Target

The developmental nature of L2 acquisition means that L2 speakers can be situated along a linguistic continuum (see a similar proposal by Polinsky and Kagan, 2007 for Heritage Speaker Grammars which represents different stages of acquisition and proficiencies). ILG are representations of specific points in the process of acquiring a second language as learners move closer to an end point. Since native controls are speakers who have a complete (or end state) grammar (see Meisel, 2011), it is legitimate to regard a NS grammar as a possible end point (target) in the L2 acquisition continuum. Thus, comparisons between the current state and the target (the end state grammar) as well as the current state and (a possible) next state (i.e., NNS-NNS comparisons) are necessary in order to understand the fundamental nature of ILG and L2 acquisition. Without such comparison, the data can be described but both meaningful analyses and predictions for subsequent development are virtually impossible.

Native Controls are Necessary in Experimental SLA to Validate the Tasks

Evidence from the behaviour of native controls is key as it ensures that the instruments are appropriate and that the theoretical assumptions are correct. We advocate for the elicitation of data from a variety of tasks so any conclusions on the nature of ILG are based on more than one source of evidence. Native speakers are not chosen as the baseline because they are perfect, privileged or infallible but because they are often the control group that is methodologically appropriate. This is why native speakers undergoing attrition are not appropriate controls for SLA studies (their grammars do not represent the end state of language acquisition anymore), although they may be appropriate controls in other contexts.

The Control Group Needs to Be Appropriate for Each Specific Study

The control group and the experimental group need to be matched on a number of variables to ensure that they differ only with respect to the condition to be investigated. Since some variation in the behaviour of NS is expected, it is essential that both groups speak the same variety (i.e., be exposed to the same evidence available in the input) and have the same level of literacy (same educational background) and same background characteristics (see Dąbrowska, 2010; Hulstijn, 2011; Andringa et al., 2012; Hulstijn and Andringa, 2014). The challenge is to decide what group of NS to include for the comparison with NNS to be appropriate.

Debates on the usefulness of native controls go beyond the legitimacy of the *CF* as many believe that comparisons between native speakers and non-native speakers are *unfair* (on moral grounds) as learners/bilingual speakers are expected to conform to native norms unfairly. This is particularly critical in the case of learners of English as this language carries added connotations of colonisation, power and privilege, notions which are not the main concern of most SLA researchers. It is important to note that the original formulation of the *CF* discusses comparisons between grammars (interlanguage systems) without specifically referring to the speakers of the target language (TL) as native speakers. However, one main objection clearly concerns the use of native speakers. There are various reasons for this, one of them being that 'native speaker' carries negative connotations outside the strict SLA remit. In particular, concepts such as NS and NNS are used to represent the people themselves, even though for linguistic-cognitive SLA approaches (as well as for Bley-Vroman), the focus is on the linguistic system, not the speakers in their social context.

One consequence of the misunderstanding of what the object of study is (grammars vs. people) is that any analysis or

[2]Tensions between the so-called cognitive approaches to SLA (those which investigate linguistic systems) and sociocultural perspectives (those which focus on the socialisation aspect of language learning) are not new (see Zuengler and Cole, 2005 for a review). A good example of the types of criticism that cognitive SLA has endured over the years can be found in the arguments put forward in Firth and Wagner (1997) and the defence in Gass (1998) and Long (1997). We see recent criticisms of the type expressed in Ortega (2014, 2019) as another turn of the screw in the quest for dismissing any serious inquiry into second-language acquisition which has a theoretical interest and focuses on investigating grammatical knowledge.

evaluation in deficit/error terms can erroneously be extended to the speaker as a person. In turn, this can be used to claim that SLA researchers think of learners as being deficient speakers (Ortega, 2014; The Douglas Fir Group, 2016). Although issues around native prestige have been debated in related fields for some time [see, e.g., discussions on the superior native speaker (Phillipson, 1992)] and native-speakerism in English Language Teaching (Holliday, 2006, 2015), these are now emerging within our field. The extreme position goes as far as arguing that grammars (or ILG) are not a legitimate object of study (see Ortega, 2019), a claim which is neither in the spirit of the original formulation of the *CF* nor does it fit within the main goals of our field.

Although we argue against using the *CF* to make methodological decisions in our research, we also recognise that those working on formal/cognitive SLA approaches should pay attention to the terminology employed and the rationale for including comparisons with native speakers when this is the appropriate choice.[3] For instance, referring to 'NS-NNs comparisons' may no longer be completely appropriate in certain contexts as this is likely to be interpreted to mean that it is the speakers themselves who are being compared. It has to be clear that we are talking about systems, grammars, interlanguages and abstract grammatical representations. For that reason, we propose that in certain contexts, 'end state grammars' instead of 'native speaker' can be useful to avoid this type of confusion.

Being mindful of how we make our research findings available to non-specialist audiences is also important (see discussion in Domínguez et al., 2019), in particular when discussing notions, such as 'deficits', 'incomplete acquisition' and 'not target-like', which can be easily misconstrued. Criticisms based on the *CF* and the monolingual bias have increased at a time when SLA researchers working on theoretical issues are urged to share their research findings with people who are not familiar with our goals and methods, including researchers in other disciplines, the general public, funding agencies and the learners/bilingual speakers themselves. We recognise the difficulty in explaining notions such as deficits and errors, incompleteness and underproduction, but rather than rejecting the use of NS we ask to engage in a debate on how the field can overcome this challenge.

DESCRIBING, ANALYSING AND EXPLAINING L2 GRAMMARS

When the *CF* was proposed in 1983, the field of SLA looked quite different to what it does today, both in terms of its goals and methodology. At the time, there was an interest in investigating the systematicity of interlanguage grammars (ILG), the oral language produced by L2 speakers (Nemser, 1971; Selinker, 1972).[4] One of the main assumptions of interlanguage studies, inspired by generative studies (Sorace, 1996), is that ILG are systems governed by rules. This implies that ILG are systematic, although some variation in the behaviour of learners is expected as well (see Andersen, 1977; Hyltenstam, 1977; Dittmar and Klein, 1979; Tarone, 1979; Meisel et al., 1981; Clahsen et al., 1983; Ellis, 1985; Schachter, 1986).[5]

Brey Vroman's (BV) rationale for proposing the *CF* was based on his criticism against how systematicity was being investigated at that time by studies using oral production data as evidence. BV's focus is on how researchers can best describe the ILG without involving the target language. In interlanguage studies, grammars are systematic if they conform to certain rules and expectations which need to be established by the researcher and are based on analyses of the target language (TL). How can the researcher know what the learner is thinking or what the 'internal logic' of the ILG is? For BV, this question cannot be answered with the analytical tools employed at that time, mainly searching for contexts in which a specific form *should* be used (the so-called obligatory context). Pica (1983,p. 70) explains that '*Suppliance in Obligatory Contexts (hereafter also SOC) is used to determine accurate suppliance of morphemes in linguistic environments in which these morphemes are required in standard English*'. The notion of SOC has been instrumental in morpheme studies (Brown, 1973; Dulay and Burt, 1974) which have focused on tracking the emergence and use of morphological forms in English. SOC was criticised at that time because it cannot reveal whether the learners have acquired all patterns and distributions of use of the target forms[6]. As Pica (1983) agues, target-like use analysis (TUA) can provide this insight as it also includes the number of non-obligatory contexts in which the target form is supplied inappropriately. A review of these two analyses by Pica (1983), however, shows that when applied to the same data set, they render different results so different interpretations can be made depending on how the target forms are quantified. The point that BV is trying to make, however, is that SOC is not learner-based as it does not emerge from observations of the learner data alone but by comparisons with the target grammar. In particular, he criticises the methodological approach described in Tarone et al. (1976) as it is based on '*the mistake of studying the systematic character of one language by comparing it to another*' as well as the fact that

[3]A type of social injustice exists in Academia as English is the dominant vehicular language for transmission and dissemination of research findings. Today, most of the high-impact journals in our field overtly or covertly support a policy by which authors must comply with English native-like writing form and style (see Costello, 2020 for the privilege of English in academic publishing).

[4]For an overview of some main interlanguage studies, see Han and Tarone (2014) and Tarone (2018).

[5]It is important to note that Selinker (1972) originally proposed that ILG can only be studied when the learner engages in oral communication, a view which was not shared by Corder (1981) who argued in favour of investigating the learner's judgment on grammaticality as well.

[6]Methodological issues in morpheme studies have also been raised by Andersen (1977), Hakuta (1976), Hatch and Wagner-Gough (1976), Lightbown et al. (1980) and Stauble (1981).

'*it obscures the internal structure of the learner's system (p.6)*'. This is the Comparative Fallacy (*CF*).[7]

BV points out a number of problems with the methodology that Tarone et al. (1976) used to investigate systematicity, including that they cannot discern subcases of obligatory context and that the applicability of their measure is unknowable because one cannot tell whether the learner is faced with a binary choice as they assume. He argues that the binary nature used in SOC studies lumps together many possible options which the learners may have entertained but which cannot be revealed by the limited nature of the options made available by the researcher. He also notes that the linguistic analysis that the researcher brings to the ILG may not be available to the learners. This implies that the mere speculation of an obligatory context is a case of a comparative fallacy. Since the description of the ILG has to be done independent of the TL, the use of obligatory context (or any assumption that a certain form should be used) is, indeed, discouraged as well. If there is no possibility of any expectation of use of a form, then other key notions, such as accuracy or errors, should not be used either. This, in turn, implies that even describing whether learners use a form using percentages (e.g., reporting that a certain form is used an x number of times), which is common practice in the field, has to be abandoned too. The point is that adopting the *CF* has a knock-on effect on the whole range of methodological choices and types of analyses available to researchers well beyond NS-NSS comparisons.

Other concepts and tools that should not be used for the same logic are proficiency scales (beginners, intermediate, advanced, near-native etc), omissions, overproductions, simplifications and all of the other typical characterisations of interlanguage grammars [for an example of an analysis without these concepts see Klein's (1998) description of the Basic Variety]. In fact, even investigating whether a form or structure has been acquired is a case of the *CF* as this question already imposes a view of the learner system based on what is observed on the TL and not their internal logic. It is clear that adopting BV's own interpretation that ILGs are systems in their own right is at odds with one of the main assumptions in our field: that L2 speakers are learners engaged in the process of learning the grammar of a second language and that in this process they entertain different linguistic systems until they reach the end state (the target grammar).

For this reason, it is important to understand that adopting the *CF* has important methodological consequences involving the tools that researchers can or cannot use in their research. It is often the case that researchers who choose not to use comparisons with native controls still analyse the learner data in terms of accuracy and expected use/acceptance of forms, even those this necessarily assumes the existence of a baseline and, thus, promotes the *CF*. For instance, Schwartz (1997) agrees that UG-oriented SLA suffers from the comparative fallacy because the ILG is judged against norms from the target language. However, she also claims that '*From this perspective, that properties of the TL do not get acquired requires explanation*'. Implicitly, Schwartz still assumes that L2 acquisition involves acquiring features present in the grammar of another group of speakers who are not the learners (i.e., absence of a required feature is an error). Schwartz's (1997) view, with which we agree, still constitutes a case of the comparative fallacy according to Bley-Vroman's own definition.

Furthermore, by focusing on descriptions of ILG only, BV avoids the fact that his proposed methodology makes it virtually impossible to provide meaningful explanations about the nature of ILG and the process of acquiring a second language. Thus, the main problem arising from adopting the *CF* is that it does not fulfil the explanatory goal of the field. At the time when BV proposed the *CF*, the focus may have been on providing descriptions of ILG but this does not meet the main goals of the field[8] today which include to (1) describe (2) analyse and (3) explain the process of acquiring grammatical systems (see Gass, 1998; Norris and Ortega, 2003). Adopting the *CF* in its strong form is problematic as researchers could only (1) describe ILG but not (2) analyse or (3) explain (evaluate on theoretical grounds) the evidence. This position strongly contradicts the main goals of the discipline as stated above. A soft version of the comparative fallacy is also possible: the only goal of SLA is to (1) describe and (2) analyse ILG avoiding (3) to explain (i.e., evaluate on theoretical grounds) the evidence. This version also explicitly excludes comparisons with controls, and it is in line with the original spirit of the *CF* (i.e., to provide the right kind of descriptions emerging from the learner data only). However, it is also in contradiction with the main goal of our field as it necessarily precludes an interpretation and evaluation of any finding. For instance, if a group of learners are found to use the definite article in some contexts (a description of the data without quantifying the use by means of an obligatory context), we would not be able to interpret this finding to be low or high if we do not know what the expected use is as set by speakers who already have that form in their grammatical systems. The only way that research can provide meaningful and appropriate analyses of ILG and test hypotheses which investigate the acquisition process is by comparing ILG with the target grammar.

The Comparative Logic is the only position that can achieve the three goals of SLA research: (1) describe (2) analyse and (3) explain ILG. This position justifies the use of controls and comparisons between grammars from learners and a baseline on purely scientific grounds. The baseline for L2 studies is often formed by native speakers but this is not necessarily always the case (e.g., two groups of learners to investigate L3 acquisition; comparing second- vs. third-generation bilingual

[7]We note that although there is no explicit mention of the native language by Bley-Vroman, in principle, comparisons between ILG and any other language (native or not) would also fall foul of the *CF*. This would include comparing L2 and L3 speakers or two groups of bilingual speakers (e.g., second vs. third generation heritage speakers) or even comparing the same group of learners at different points in time in a longitudinal study. The specific objection to imposing native norms on learner grammars, although related, is in fact independent of the *CF* even though they are usually interconnected.

[8]See Zuengler and Miller (2006) for a discussion on the main opposing perspectives in SLA research focusing on the long-standing debate between cognitive and sociocultural approaches.

heritage speakers; comparing the native language of monolingual and bilingual speakers undergoing attrition). As we will argue in the following sections, native speakers become legitimate members of a control group because of the nature of their grammatical systems, not because they are ideal or infallible speakers. It is also possible that certain grammatical areas may be subjected to a higher level of variation than others even for native speakers. This is why it is informative to collect these data from a control group in experimental SLA studies.

In summary, in this section, we have shown two main problems with adopting the *CF*; first, researchers lose the main methodological tools and concepts which are necessary to analyse and explain learner grammars (error, accuracy, overproduction, etc); second, the possibility of providing meaningful explanations is virtually impossible if there is no link between the learner and the target grammars. We have argued that comparisons, including NS-NNS comparisons, are necessary to meet the main SLA goals, a position which we have called the Comparative Logic.

DEVELOPMENT IN SLA: ACCEPTING THAT L2 SPEAKERS ARE LEARNERS

The view that ILG are systems in their own right, which can be traced back to at least Selinker's (1972) original definition of interlanguage, is widely adopted in our field. Bley-Vroman agrees with this view as well but also argues that ILGs need to be analysed independently of any other system as this is the only way that the own logic of ILG can be revealed; for this reason, he claims that the comparison with the grammar of speakers of the target language (TL) makes ILGs degenerate versions of the native grammar. It is important to note that the word chosen by BV is 'degenerate' which means degraded, abnormal and of lower quality. In our view, degenerate is an unfortunate choice of term as it is a measure of quality (i.e., non-native speakers produce language of substandard quality) which does not naturally arise from the objective description of that system.

Some researchers have taken the view that if L2 grammar lacks a grammatical feature or contains an error, that means that the speakers themselves are deficient in some way (see Firth and Wagner, 1997).[9] Although this misconception has been already addressed by some (see Gass, 1998), criticisms of this kind towards linguistic-cognitive SLA research still remain (Ortega, 2014, 2019). Reconciliation on this matter necessarily entails an understanding of how 'deficit' is understood in linguistic-cognitive SLA and why it is important that we investigate both what learners can and cannot do in the process of acquiring a second language. Although a deficit view of acquisition (both for first and second-language acquisition) exists, this is to mean that learners make errors or show incomplete knowledge of a certain grammatical aspect of the TL, not that the learners themselves are deficient in any way. Both fossilisation (Selinker, 1972; Han, 2004) and incompleteness (Schachter, 1988, 1990; Sorace, 1993) have been routinely used to describe aspects of learner grammars. These terms only make sense because ILGs are evaluated against a target (complete) grammar where target means that it represents the outcome of language acquisition under ideal input conditions (what we will characterise in the next section as the 'end state'). We have already argued that since the *CF* prevents researchers from making any evaluations of ILG that would conclude that the system is degenerate (incomplete or deficient), concepts which are widely used in our field, such as errors, omissions, overgeneralisations and simplifications, would need to be abandoned as well. In our view, this is the wrong approach as we would stop using the tools that allow researchers to carry out explanatory research in second-language acquisition. For this reason, it is our view that any research committed to offer precise descriptions and explanatory answers will necessarily be subject to, at least, the soft version of the comparative fallacy as any explanation arising from the description of ILG would necessarily need to address the deficit/error issue we just noted.[10]

In fact, adopting the idea that the *CF* exists intrinsically threats the notion of interlanguage itself, as interlanguage was proposed as a means to account for the process involved (often shown by different stages) in learning a target language. In traditional interlanguage studies, assuming that learners develop a second language (i.e., they move closer to the TL) does not necessarily mean that an ILG is not a system in its own right but, rather, that the learner is in the process of acquiring a full grammatical system with all the features expected in that system. For instance, Spanish has grammatical gender which triggers a type of agreement between nouns, adjectives and determiners (e.g., *la gata negra*/the black female cat). Thus, it is reasonable to expect that learners of Spanish will have to learn this feature which is likely not to be present at the early stages of acquisition. Until that feature is present in their grammar, the process of acquiring Spanish (the target grammar) can be said to be incomplete. Researchers interested in finding out how learners go about the challenge of acquiring a new feature (gender) which does not exist in their native language need to know whether learners use the right gender (masculine or feminine) appropriately. It would not be possible to do this without a reference to how gender is used by speakers of the target language.

In this respect, one basic assumption in SLA studies is that we are investigating a process whereby a speaker *develops* a second/n language through a specific route. An ILG represents specific points in the process of acquiring a second language (see Meisel et al., 1981 for a discussion on developmental stages in L2 acquisition). This process necessarily entails a progression which, in turn, necessarily assumes that certain features of the target grammar can/

[9]For a similar argument in the context of heritage language acquisition, see Domínguez et al. (2019).

[10]As an example, the Basic Variety was proposed with this objection in mind and it is the result of a description of a learner system without references to a TL. This approach is rather limited in its explanatory power and the kinds of predictions it can make about the SLA process more generally.

should/will be absent. In this respect, one can argue that this represents a 'deficit' view of language development, similar to what is observed in the process of acquiring a native grammar in the case of children, as a learner starts the process of acquiring a language with very little knowledge of the grammar which is being learnt. Deficit in this context does not mean that the grammar of a learner is of lower quality, degraded or degenerate, nor is it an evaluation of the speakers themselves. It means that the system entertained by the learner does not (yet) show the features and properties of the target grammar. Importantly, this 'deficient' or 'incomplete' view is not in opposition to the view that learner grammars are systems in their own right.[11] Both interpretations can be true. This point becomes clear when analysing overregularisations, such as when learners use the English past tense marker *-ed* with an irregular verb (e.g., using 'goed' instead of 'went'). The use of 'goed' is both an error (i.e., it is not how the verb 'go' marks past tense in English) and the result of the speakers' grammar respecting a certain grammatical principle of their own system (e.g., use 'ed' with all verbs to mark past tense).

We would like to reiterate our point that without comparisons with the TL, there can be no analysis, and that without analysis, there cannot be any explanations. Notions, such as accuracy and errors, are fundamental to understand the processing of acquiring a language in all contexts. There are numerous examples of how different SLA frameworks make notions, such as accuracy and error central to their analyses. Without these, there would be no field. A good overview of some ways in which interactionist, emergentist and generative scholars measure SLA is found in Norris and Ortega (2003). For instance, these authors show that detecting the use of a form is important for interactionist approaches to SLA. However, this is not the measure use for acquisition as learners have to show that they are also able to use that form *appropriately* and fluently. The only way in which it makes sense to describe the use of a form as appropriate is if some criteria for such use has been established.

With regard to emergentist approaches, Norris and Ortega (2003, p. 727–728) explain that accuracy is one of the main factors used for establishing the parameters of acquisition in this framework. As in the case of appropriateness above, accuracy can only be established if a comparison with a 'correct' use of the form is established. In these two frameworks, comparisons between learner grammars and the grammar of the TL are necessary to fulfil our goal of understanding the process underlying SLA. Interestingly, the term 'nativelike' is only mentioned by Norris and Ortega when they describe generative approaches to SLA: '*Generative linguistic studies of SLA are likely to rely almost exclusively on the outcomes of grammaticality judgment tasks of various kinds, where acquired means nativelike levels of rejection of illegal exemplars of the target grammar*'. Although we agree with these authors that the term nativelike is often used by generative SLA studies, the same concern with the appropriate and accurate use of target form is shared by all of the frameworks reviewed by these authors. For all these researchers, the use of target forms is analysed by comparison with a group of speakers which perform target-like. That is, one fundamental notion of acquisition is that it assumes conformity with native use/judgement in all approaches. For instance, in a study promoting task-based learning, Pica et al. (2006, p. 320) describe ILG as being full of omissions, substitutions and inconsistencies and a varying degree of accuracy. They do this without explicitly comparing learner behaviour with a group of native control even though this is the only one in which they can discuss accuracy. An important body of research has been concerned with the role of corrective feedback in SLA. Studies on corrective feedback assume that L2 learners make errors. For instance, Ellis et al. (2006, p. 340) argue that '*Corrective feedback takes the form of responses to learner utterances that contain an error. The responses can consist of (a) an indication that an error has been committed, (b) provision of the correct target language form, or (c) metalinguistic information about the nature of the error, or any combination of these*'. Superficially, one could conclude that the focus of the investigation is to show that learners fail to acquire a second language. Similarly, approaches which investigate NS and NNS interactions (see, e.g., Lyster and Saito, 2010) do so on the assumption that the NS plays a crucial role in second-language development: it is through the interaction with a NS that input is rendered comprehensible to learners. Finally, when Andersen and Shirai (1994, p. 143) proposed the extremely influential 'Aspect Hypothesis' to explain the L2 acquisition of past tense morphology, they were trying to explain why learners fail to supply past marking in obligatory context much more frequently with some predicates than with others. The analysis of correct and incorrect compliance of target forms was the basis of Andersen and Shirai's analysis later adopted by a large number of studies.

These examples show how in all of these approaches, notions such as accuracy, progress and errors, are crucial if it is expected that ILG develops towards a target. As Lardiere (2003) argues, even those approaches/researchers who are supposed to be respectful of the comparative fallacy (because they claim that they investigate learners' interlanguage on its own right) are susceptible of it once they base their analysis on notions, such as obligatory context, accuracy and omissions. In our view, understanding and explaining SLA necessarily require comparisons with a baseline. We have called this the Comparative Logic and have argued that it is the most appropriate position in order to both view ILG as system in their own right and provide meaningful explanations. Analysing and understanding when success is both possible and when it is fundamental in our field.

[11]One obvious consequence arising from the developing nature of ILG is that learners are often classified according to different proficiencies. This classification assumes a comparison as well, not between learners and native speakers, but between learners and other more or less advanced learners. Since a beginner learner is only a beginner in comparison with a more proficient learner, this type of comparison should also present a case of the *CF* as per BV's definition. This means that across-group comparisons between learners (not just native speakers) are to be avoided as well.

GENERATIVE SLA AND THE ROLE OF NATIVE CONTROLS

In the previous section we showed that comparisons between learner and complete (native) grammars are commonplace in the field because they are necessary to explain ILG irrespective of the theoretical framework; however, it is often the case that researches in the generative tradition are the target of criticism specifically for promoting comparisons between learner and native grammars and for the (erroneous) belief that the field sees native speaker norms as a goal. This is partly due to the fact that having evidence from native speakers' intuitions is clearly part of the methodological design. There are other reasons which are linked to the main assumptions of the whole generative enterprise which have been carried over to SLA research. As we have already explained, generative SLA is concerned with the abstract linguistic knowledge of speakers, what they unconsciously know about language(s). The field assumes an innate and biologically determined capacity for language which is unique to humans. The specialised and abstract module specific to language known as the computational system includes a lexicon and the syntactic operation Merge (Chomsky, 1998; Berwick and Chomsky, 2008; Friederici, 2017) which builds syntactic structures which are interpreted and pronounced by specific subsystems. Importantly, there is evidence suggesting that access to this capacity may decline with age as differences between how speakers acquire a native and a non-native language have been found (see discussion in White and Genesee, 1996).

Unlike other cognitive approaches, generative SLA is interested in I-language, rather than language as a social or cultural object. I-language is an internalised system, what is also known as a grammar. I-language is according to Chomsky et al. (2019) '*a system that links meaning and sound/sign in a systematic fashion, equipping the speaker with knowledge of these correlations*'. During the language acquisition process, assumed to be constrained by Universal Grammar (UG), children develop a grammar (i.e., they figure out what is correct and what is not) and establish form and meaning pairs as determined by the language faculty (Chomsky, 1986). These form-meaning connections thus exist in the target language which serves as the input for L2 speakers. Typically, the language acquisition process finishes when children's grammars reach the so-called 'steady' or 'end state'. The 'steady state' is the full adult grammar resulting from full access to UG and exposure to a full set of linguistic input; in this respect, one could say that it is what results in 'ideal conditions' for language acquisition in the sense that full convergence with the 'end state' is always achieved. For this reason, we argue that a more appropriate way of calling native speakers in SLA research would be 'end state speakers' or even more appropriate those who have an 'end state grammar' to avoid any confusion about what the object of our study is.

In the context of L2 speakers, ILG is also an I-language (see Adjemian, 1976; Klein, 1998). L2 speakers have access to UG[12] during the acquisition process but the characteristics of their 'steady state', unlike the case of children, are unclear. It is also not completely obvious whether any intermediate grammars or ILG have direct access to UG or whether all L2 speakers reach a similar 'steady state' with the same characteristics.[13] Comparing the status and characteristics of these intermediate I-languages and the corresponding 'end states' is useful to evaluate the role and accessibility of UG, the role of the input during acquisition, L1 influence, etc. Even though updated views of the role of UG have promoted other types of research questions (the role of linguistic interfaces, representational impairment vs. computational efficiency, feature-reassembly etc.), White's (2003) claim that '*the crucial question is whether or not interlanguage grammars are UG-constrained, rather than whether or not they are native-like*' is still valid today.

One specific and very common criticism against generative SLA is based on the (misinformed) claim that generative approaches to language are based on native speakers are idealised speakers (Leung et al., 1997); embedded in this criticism is, again, that speaker here refers to the speaker as a person functioning in the real world, not their abstract linguistic system as we have just explained. This particular criticism often arises from a misunderstanding of what 'ideal' means[14] and the reasons that led Chomsky to propose this assumption in the first place. The contentious quote from Chomsky (1965, p. 3) is as follows: '*Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance*'. This may appear to be a call for a search of the perfect speaker, which is identified with a native speaker (i.e., nativism equals perfection). However, Chomsky is really arguing that in order to understand grammar as a cognitive system (competence), one has to look further than what speakers actually say (performance) as this is modulated by non-linguistic factors. Chomsky is concerned with knowledge of a grammar as an abstraction, an outcome

[12]Whether UG is available for L2 acquisition has been a major topic of debate (see Borer, 1996; Epstein et al., 1996), the current view being that learners can indeed access UG when they encounter input which cannot be comprehended/parsed by their existing grammar (the Full Access position; see White, 2003).
[13]Bley-Vroman (1990, 2009) argues that the process of acquiring a first and second (foreign) language are fundamentally different. The Fundamental Difference Hypothesis (FDH) proposes that whereas child language acquisition may be constrained by UG (which is domain-specific), foreign language learning is not, and so adults need to resort to general learning mechanisms in an instance of general skill acquisition. He makes this claim on the basis of certain observations which have mostly been contested or not completely accepted including that 'complete success is extremely rare, or perhaps even non-existent' (1990:6); adults not only generally do not succeed, they also fail to different degrees (1990:7); and adults set their own goals and can fail, this leads to variation in the process and outcomes. It is interesting that despite the fact that the FDH and the Comparative Fallacy are based on a clear anti UG position, some UG scholars seem to embrace both.
[14]It is only by accident that the 'idealised' speaker coincides with the native speaker. Some native speakers are not appropriate as control groups in L2 studies, namely, those undergoing grammatical attrition (see Domínguez, 2013).

of language acquisition in ideal learning conditions, not as a real object that can be studied. In this respect, Chomsky (1965) also explains that '*a generative grammar is not a model for a speaker or a hearer. It attempts to characterize in the most neutral possible terms the knowledge of the language that provides the basis for actual use of language by a speaker hearer*'.[15] Critics also ignore the fact that Chomsky later abandoned the competence-performance distinction (and the idealised speaker) in favour of I-language and E-language (Chomsky, 1986) as this distinction, among other reasons, can account for linguistic variation.

There have been some attempts to deconstruct and even rethink the need to assume an ideal speaker/hearer both in formal and experimental contexts. For instance, Chesi and Moro (2015) discuss the competence vs. performance distinction proposing that there is both an idealised native speaker and a real native speaker. The native speakers are not idealised speakers themselves but have access to the grammar which is the object of study. This distinction is also useful for explaining why the behaviour of native controls does not always agree with the predictions made by linguistic theory (which are based on the most idealised competence systems). Similarly, in the SLA literature, Duffield (2003) distinguishes between two types of linguistic competence (underlying and surface competence) to account for knowledge of gradient grammaticality (when a structure is more acceptable than another). Underlying competence is categorical, whereas surface competence is more probabilistic as it includes several factors, such as sensitivity to frequency of constructions. More recently, Slabakova et al. (2011) have provided empirical support for Duffield's dual competence system and Sorace and Keller (2005) have also made a similar distinction between hard (syntactic) and soft (interface-based) constraints which yield different levels of acceptability.[16]

The use of native speakers as native controls has been often justified on methodological grounds. Research which focuses on judgement data as the main source of evidence requires a control group in the experimental design. Since the control group is very often a group of native speakers of the TL, although not exclusively, the comparison between native and non-native behaviour is often made explicitly. As Sorace (1996, p. 380) notes: '*For the correctness of judgments to be empirically assessable, it should be possible to measure intuitions of degree of grammaticality against some independently established grammaticality scale*'. Sorace's quote shows that comparisons among groups (including a baseline of native speakers) are necessary for explaining and assessing the results arising from linguistic judgements. It is an essential part of the experimental design used in research which investigates learners' judgements and intuitions. In this type of experiments, a set of variables are defined and controlled. The control-experimental group comparison is also necessary to determine whether the results are the effect of the independent variables or not, to establish the baseline of comparison, to verify the validity of the task and for investigating whether the hypotheses are incorrect and need to be reformulated. There are numerous examples showing that this has been the case in Generative SLA. For instance, Grüter (2006) in a key study which found support for the Full Transfer/Full Access position used a control group to analyse the acquisition of wh-questions in German. The behaviour of the control group is key to show that there is a bias for one of the two possible readings of a question which was not expected nor found in the L2 data. Without the native control data, some of the learner behaviour would have not been explained by the hypotheses.

If native speakers are necessary as baselines to control conditions and offer a key measure for understanding learner behaviour, how can researchers meet their methodological needs and avoid the *CF* at the same time?[17] This is definitely a challenge for UG-based research which has an explanatory goal that goes beyond providing descriptions and often elicits intuitions; in fact, such is the difficulty that we argue that it is virtually impossible. In our view, the key is to separate that comparisons between native and non-native grammars are necessary from any conclusions that researchers can reach based on those comparisons (the issue is how 'deficit' is/should be approached). In particular, it should be possible to investigate learner grammars in their own right while providing analyses which take into account the judgements of speakers of the target language. In this respect, we agree with Sorace (1996, p. 385) that even when the comparison with native speakers are justified '*learners' judgments themselves should provide the primary criterion for deciding which structures are or are not part of it (the non-native grammar)*'. The learner's data are still the relevant data as argued by Birdsong (1989).

Those researchers which still choose not to include a group of native speakers as control groups need to clearly specify how they intend to provide accurate and appropriate descriptions and explanations of the learner data. For instance, Heil and López (2020) included a group of native speakers as controls but learners' and natives' judgements were not analysed together. The authors showed the results of the monolingual English

[15]Traditionally, generative SLA research is often difficult to translate and apply to the real world (as opposed for instance to pedagogy-oriented SLA). As an example, it is not immediately obvious how examining UG accessibility can be of any use to foreign language teachers or even to the learners themselves.

[16]It is generally agreed that grammaticality taps into speaker's competence (or I-language) and is not open to gradience whereas acceptability involve the speakers' performance and is gradient in nature. This is because speakers are judging sentences according to their perception of those sentences (Bard et al., 1996). For a full discussion on the differences between grammaticality and acceptability plus how the parser can also affect acceptability, see Leivada and Westergaard (2020). For a detailed discussion on how the interpretation of gradient judgments affects sentence acceptability, see Francis (2021).

[17]One anonymous reviewer suggests that we consider the suitability of using native speakers who are also L2 learners as controls in SLA studies on the grounds that both groups would be bilingual. It is our view that the characteristics of the control group depend on the research question to be investigated and so studies who are interested in investigating the effects of the bilingual experience should take this variable into account when selecting the controls. In some studies having two control groups (one formed by monolingual speakers only and one formed by monolingual speakers who also know another language) may even be relevant. The reviewer's suggestion is consistent with our view that the selection of the control group should be carefully considered for each particular study.

group in order to verifying experimental validity as they wanted to avoid the *CF*. In the method they use, they provide evaluations of learners' grammars based on indirect comparisons with native controls. However, it is difficult not to draw comparisons between these two groups when both sets of results are presented together in the same tables and there is a clear connection between the behaviour of the learners and the native speakers. Furthermore, there are studies in which a direct statistical comparison between the control and the experimental groups is justified. This comparison, which is essential in certain studies, should not be ruled out on the basis that it provides a case of the *CF*.

VARIABILITY IN THE (NATIVE) CONTROL DATA

One final argument against NS-NNs comparisons is that the NS themselves do not form a homogeneous group and variability in the data makes it difficult to set goals for learners based on how we expect NS to behave. In this section, we argue that variation within a community of speakers and within speakers themselves is nothing unusual and has been successfully accounted for in linguistic theory. We will also show how some of the concerns raised with respect of variability can be mitigated by applying more rigorous research methods, in particular better sampling techniques.

Formal SLA has borrowed analytical tools from linguistic theory as researchers assume that evidence of knowledge of grammar is shown by knowing what is both grammatical and ungrammatical. There is also a long tradition of testing hypotheses in controlled, experimental settings.[18] *A priori* it may seem that variation is problematic for a UG approach to language since UG is invariant by nature. However, variability has been accounted for by several approaches, such as Adger's (2006) Combinatorial Variability model or the Multiple grammars approach (Kroch, 1989, 1994; Yang, 2002). Another recent development has brought together generative syntax and variationist sociolinguistics [see review in Adger et al. (submitted)] and employs a new methodological approach which moves beyond the individual and focuses on both linguistic and social aspects of the whole community of speakers. Under this approach and following Labov (1982), it is expected that the linguistic rules shared in the community are of a variable nature. Sentences which would be ungrammatical for some speakers of English can be part of the grammar of speakers of certain varieties for which the standard and the regional variety are both possible. For instance, Henry (1996) shows cases of word order variation with imperatives in Northern Irish English as shown in (1a) and (1b):

1.
   a. You go away
   b. Go you away

Despite the fact that intra-speaker variation is often observed as shown in example (1), there is still an expectation that speakers would conform to certain rules, that is certain aspects of the grammar are not subjected to intra-speaker variation regardless of differences in gender, class, style, education, age etc. For instance, sentence (2b) with a missing subject would not be acceptable by any speaker of English:

2.
   a. Lena says that [she] will come soon
   b. *Lena says that [] will come soon

There is some tension between conformity and variability when investigating the linguistic behaviour of speakers. We expect speakers of English to conform to core syntactic properties (such as the use of overt/null subjects) in some cases more clearly than others. It is important to highlight that cases, such as the examples shown in (1) are cases of *true* variability in the speakers' grammars (I-language). However, the SLA literature also describes a type of variability which is linked to performance and to other methodologically related issues. For instance, Sorace (1996, p. 377–378) mentions several extralinguistic factors that are likely to influence how participants go about completing grammaticality judgements including parsing strategies, context and mode of presentation, pragmatic considerations, mental states and linguistic training. Schütze (1996/2016) also shows that literacy is a relevant factor. These and other similar factors, which are external to the mental representation of the grammar, are important for SLA researchers and can affect the results arising from grammaticality/acceptability tasks giving raise to extralinguistic variation. Researchers should try to minimise this by choosing the appropriate design and research method.

In particular, it is important that for some structures, researchers allow for the possibility of using gradience or a range of responses (usually a Likert scale) rather than restricting the responses to yes/no answers (see discussion in Schütze, 1996/2016). In some cases, it may be necessary to elicit evidence through various types of tasks and make comparisons based on a range of answers rather than a fixed point (see, e.g., Hyltenstam and Abrahamsson, 2000; Abrahamsson and Hyltenstam, 2009) and how they judged the performance of L2 speakers against the whole range of responses provided by native speakers). Recruiting participants to be part of the control group is an important task which needs careful attention from the part of the researcher (see Lipsey, 1990; Quené, 2010) so that the sample is both as homogeneous and representative

[18]In theoretical syntax the research method is to obtain a judgement of the acceptability of a sentence often by just using the intuition of the author/s of the study. Phillips (2009) and Adger (2014) have both defended this method. Experimental evidence supporting the validity of introspective judgements can be found in Sprouse et al. (2013) as they show that these data have medium to large effect sizes. Sprouse (2011) argues that the results of an acceptability judgement task conducted *via* Amazon.com's Mechanical Turk (AMT) are almost indistinguishable from the results from the same task conducted in a laboratory with control from the researcher. AMT has the key advantage that it can recruit participants for the control group from a wide pool and varied backgrounds. In contrast, Gibson and Fedorenko (2013) point out some of the problems including that this method leads to incorrect generalisations due to cognitive biases from the part of the researcher. They argue for a quantificational approach using Amazon.com's Mechanical Turk (AMT).

as possible. Special attention needs to be paid so that the control group and the experimental group are matched on the key variables to ensure that they differ in respect of the condition to be investigated only. Other adjustments, such as that both groups speak the same variety and are exposed to the same evidence available in the input, should be taken into account as well.

It has also been argued that other factors, such as processing and experience, may be subjected to variation. When investigating individual differences in L2 acquisition, Andringa et al. (2012) assume variation in listening proficiency for both non-native and native speakers. They found that the success comprehension process for native speakers depends on their ability to deal with the pressure of online speech processing. Those speakers with more accumulated experience processing complex texts were the best listeners. This suggest that NS should be matched with NNS of similar literacy levels. A similar argument has been made by Hulstijn and Andringa (2014) as they argue that it may not be possible to single out a single factor responsible for variation in their native control data as effects of working-memory capacity, reasoning ability and reaction-speed in a nonverbal task together could explain effects of age and length of exposure. In general, these and other studies investigating individual differences reach the conclusion that NS-NNS are legitimate as long as the right NS are included in terms of literacy, educational background, experience, background characteristics etc. Individual variation can also be an effect of the task. In this respect, Hulstijn (2011, p. 236) shows how individual differences in some tasks employed are mainly restricted to differences in the speed with which linguistic information can be processed (*as a function of age*), whereas in other tasks, it is 'mainly by differences in intellectual skills and amount of reading and writing activities, as reflected by education, occupation and leisure-time activities'.

Finally, the type of predictions and expected results can have an effect on the results as well. For instance, in our investigation of the use of preterite and imperfect forms in Spanish by both native and non-native speakers, we asked all the participants to complete a series of oral and comprehension tasks (see Domínguez et al., 2013). We investigated whether the predictions of Andersen and Shirai's (1994) Aspect Hypothesis (AH) hold for both groups so it was important to have data showing the use and acceptability of the target forms for the native speakers and the learners. According to the AH, preterite tends to be used with telic events rather than with atelic events; on the other hand, the imperfect is preferred with atelic events. The results of two oral production tasks, an interview with an investigator and a picture-based story retell show that, despite some variation in the amount of preterite and imperfect forms produced by the controls, the averages conform to the expected results. For instance, in the interview, the least-controlled task, the native controls used the preterite with achievement (telic) verbs on average 80% of the time, whereas they use this form on average 32% of the time with state (atelic) verbs. Although most native speakers used the preterit between 80 and 95% of the time with achievements, the range of use was wide from 57 to 100%. The range of use of the preterite with states was equally wide from the lowest use of 7% to the highest use of 55%. Despite this variation, the means were useful as they corroborated our predictions and showed differences with the pattern of use shown by the learners. We were able to conclude that the pattern of use of preterit and imperfect predicted by the AH is already represented in the pattern of use of these forms in the native input, so learners have access to that kind of evidence though the course of acquisition.

In Domínguez and Arche (2014), we reported variability in the data of the (native) control group even though this was not expected. All the participants completed a content-matching acceptability task to investigate preference of SV and *VS* orders with different types of verbs (accusative and unergative) and different types of pragmatic contexts (narrow focus on the subject or not). The theoretical analysis adopted predicted that native controls would prefer the *VS* structure with narrowly focused subject with unergative verbs (smoke, dance, sneeze and cry). However, the aggregated means of all the native participants showed that these speakers only chose this structure 45% of the time. A closer look at the individual results revealed that this was not a case of optionality, as native speakers had clear patterns of behaviour as roughly half of them preferred SV and roughly other half preferred *VS* in this context. Interestingly, the advanced learner group also showed variability in their responses, but in this case, the same participant would choose both options. Unlike the native controls, learners did show optionality in their responses. Based on the responses of the native control data, we were able to suggest that the input can be vague with respect to SV and *VS* structures in Spanish which can lead to difficulties (optionality) for learners.

In this section, we have argued that variation in the data is not unexpected and can be accounted for both theoretically and empirically. A more careful selection process for the control group can mitigate problems arising from extralinguistic variation and ensure that the sample is representative and appropriate.

WILL THESE ISSUES EVER BE RESOLVED? SOME REFLECTIONS FOR THE FUTURE

A review by Zuengler and Cole (2005) shows that criticisms against the goal and methodology employed by cognitive approaches to SLA have been raised for quite some time. In that review, it was clear that the criticism came from scholars from the socio-cultural tradition (e.g., Firth and Wagner, 1997). It is now the case, however, that questions on the role of the native speaker are being asked from within the cognitive field. We have analysed Bley-Vroman's Comparative Fallacy and examined the validity of its assumptions in the context of SLA research today. We have concluded that by ignoring the target grammar, the *CF* does not enable researchers to achieve the main goals of our field. This is because making methodological choices on the basis of the *CF* entails much more than not including control groups of native speakers. Those who choose

to avoid the *CF* would not be able to make any *a priori* predictions that would impose their own analysis/expectations on the learner data; they would not be able to analyse the data in terms of what is not produced, whether forms are absent or overused or simplified etc. The analysis they produce will not be able to make references to errors or accuracy either. Since these are notions which are essential to account for the nature of the acquisition process, we conclude that adopting the *CF* will prevent researchers from providing meaningful explanations.

We have also argued that the only position which ensures that the goals of the field are met (describing, analysing and explaining the process of learning a second language) is the Comparative Logic, the view that comparisons with a control group or baseline are necessary. The field, almost 30 years after BV proposed the *CF*, is well-equipped to make comparisons between learner and non-learner grammars in a way that respects the principle that ILGs are systems in their own right. Nevertheless, careful attention needs to be paid to the methodology chosen and, in particular, the sampling process for inclusion of participants in the control group or the baseline for comparison. Researchers should consider not just what types of tasks to employ but also how variation in the cognitive skills, literacy, experience etc. of the participants in the control group could lead to variability in the results.

We have also argued, as others have before us, that for cognitive SLA, errors are an important source of information when investigating the learners' mental grammars. SLA is a process by which learners entertain different interlanguages or I-languages which may not include all of the features of the target grammar until they reach the 'end' or 'steady state'. The 'steady state' is the adult grammar which results from the interaction of UG, exposure to input and certain cognitive principles during child language acquisition. Interlanguage is a type of I-language, an abstract, subconscious and internalised grammar with characteristics similar to learner grammars (ILG). For this reason, we completely agree with Gass (1998, p. 84) when she claims that the scope of inquiry of SLA is to study acquisition and so L2 speakers in this context are necessarily *learners* and not *users* of the language.[19] In the same spirit, we emphasised in this article that in order to answer relevant questions about the nature of ILG, we need to focus on the grammatical systems and not the speakers. Crucially, our enterprise does not preclude others from studying social aspects associated with learning a second language.

We are mystified that anyone could conclude that our field *promotes* native speaker norms and that there is a monolingual bias in SLA (see, e.g., Kachru 1994). We hope that this article has shown that there is no privileged status or prestige associated with the notion of a native speaker *per se* nor that native speakers are a model or inspiration for learners [see Davies (2003) for this view]. Criticisms of this sort are particularly common when generative SLA is targeted, as it is often criticised for focusing too much on correctness and the native norms. We have shown that this is due to a misunderstanding of our goals and scope of inquiry. Since the emphasis is of generative SLA is on understanding grammars (as opposed to communication or language use) and we directly judge learners' intuitions as grammatical or not, some may think that the field sees correctness as a goal when this is not clearly the case. Nevertheless, we admit that there needs to be more clarity from our part on our goals and methods, particularly when sharing our research with non-experts. In this sense, a clearer rewording of our research questions would be a step forward. For instance, generative SLA does not investigate *if an L2 speaker can become a native speaker* but rather *if an end or steady-state grammar can be attained based on partial input after the onset of the critical period*. The problem we see with this is that the latter is harder to understand and it is not as attractive as the former, particularly as researchers are under pressure to get funding, make our research impactful to non-specialists and seek collaborations with other disciplines.[20]

We believe that this is a serious issue for cognitive approaches to SLA and generative approaches in particular. Although some good attempts to made formal SLA useful to foreign language teaching exist (Whong et al., 2013; Leal and Slabakova, 2019; Rankin and Whong, 2020), a large body of our research does not have an immediate application outside the academic remit, mostly because our concerns are theoretical in nature. This may be seen as a limitation compared to other approaches, when it clearly is not, nor does it justify a radical methodological change. Without research which engages with theoretical questions, there cannot be any scientifically inspired applications. Gregg (1996, p. 75) already cautioned that L2 theories may only have intellectual value since the problems tackled are fundamentally theoretical (as opposed to practical problems). Furthermore, Newmeyer (1988) also argue that '*progress in L2 acquisition theory, as in any other scientific discipline, comes by focusing on the explanatory problem, and not by looking over one's shoulder at the possible applications*'. The apparent (lack of) immediate applicability issue has become quite real recently for researches working on theoretical issues. As pressure mounts to make our results meaningful and impactful in the real world, we make ourselves vulnerable as opportunities for misunderstanding multiply. Something as simple as proposing as a vision of SLA based on

[19]The user-learner distinction is key to understand existing opposing approaches to English language learning and teaching: whereas SLA researchers investigate how L2 speakers go about learning a language, English as a *Lingua Franca* (ELF) researchers view these speakers as users engaged in communicative practices (see Seidlhofer, 2001; Jenkins et al., 2011). We would like to emphasise that in our view, a learner of a language can also be a user of that language in other contexts and that both approaches are not mutually exclusive.

[20]We think of this situation as a paradox. The field is urged to open up and make its research impactful to others as a way to prosper and become relevant but by doing so researchers accidentally create opportunities for misunderstanding, criticism and alienation (for instance not understanding that we are interested in analysing grammars and not people in real situations). This, in turn, raises concerns about our theoretical and methodological assumptions which can make our field less prosperous and less relevant in the end.

transdisciplinarity (The Douglas Fir Group, 2016)[21] is likely to instigate even more criticism against cognitive and formal approaches to SLA as we are singled out for not taking into account the learners' social context and that they are people who function in the real world. It is in the sense that transdisciplinary in the SLA context is a trap and not a vision all researchers see as beneficial for the field (see also Han, 2016).

[21]The transdisciplinary agenda '*seeks to integrate the many layers of existing knowledge about the processes and outcomes of additional language learning by deriving coherent patterns and configurations of findings across domains* (The Douglas Fir Group, 2016, p. 20)'. The starting point of this agenda is the social-local reality of L2 learners.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LD and MA contributed to conception and design of the study. LD wrote the first draft of the manuscript. Both authors contributed to manuscript revision, read, and approved the submitted version.

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

Adger, D. (2006). Combinatorial variability. *J. Ling.* 42, 503–530. doi: 10.1017/S002222670600418X

Adger, D. (2014). Syntax. *WIREs Cognit. Sci.* 6, 131–147. doi: 10.1002/wcs.1332

Adjemian, C. (1976). On the nature of interlanguage systems. *Lang. Learn.* 26, 297–320. doi: 10.1111/j.1467-1770.1976.tb00279.x

Andersen, R. W. (1977). The impoverished state of cross-sectional morpheme acquisition/accuracy methodology. *Working Papers Biling.* 14, 49–82.

Andersen, R. W., and Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Stud. Second. Lang. Acquis.* 16, 133–156. doi: 10.1017/S0272263100012845

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., and Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Lang. Learn.* 62, 49–78. doi: 10.1111/j.1467-9922.2012.00706.x

Bard, E. G., Robertson, D., and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language* 72, 32–68. doi: 10.2307/416793

Berwick, R. C., and Chomsky, N. (2008). Poverty of the stimulus' revisited: recent challenges reconsidered. *Proc. Annual Meeting Cognit. Sci. Soc.* 30:30

Birdsong, D. (1989). *Metalinguistic Performance and Interlinguistic Competence*. United States: Springer-Verlag Publishing.

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: the case of systematicity. *Lang. Learn.* 33, 1–17. doi: 10.1111/j.1467-1770.1983.tb00983.x

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Ling. Anal.* 20, 3–49.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Stud. Second. Lang. Acquis.* 31, 175–198. doi: 10.1017/S0272263109090275

Borer, H. (1996). Access to universal grammar: the real issues. *Behav. Brain Sci.* 19, 718–720. doi: 10.1017/S0140525X00043582

Brown, R. (1973). *A First Language: The Early Stages*. London: George Allen and Unwin.

Chesi, C., and Moro, A. (2015) in *The Subtle Dependency between Competence and Performance: 50 Years Later: Reflections on Chomsky's Aspects. Vol. 50*. eds. A. J. Gallego and D. Ott (Cambridge, Mass: MIT Press), 33–45.

Chomsky, N. (1965). Persistent topics in linguistic theory. *Diogenes* 13, 13–20. doi: 10.1177/039219216501305102

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. United States: Greenwood Publishing Group.

Chomsky, N. (1998). Some observations on economy in generative grammar. *Best Good Enough* 1, 115–127.

Chomsky, N., Gallego, Á. J., and Ott, D. (2019). Generative grammar and the faculty of language: insights, questions, and challenges. *Catalan J. Ling.*, 229–261. doi: 10.5565/rev/catjl.288

Clahsen, H., Meisel, J. M., and Pienemann, M. (1983). *Deutsch als Zweitsprache: Der Spracherwerb ausländischer Arbeiter. Vol. 3*. Netherlands: John Benjamins Publishing Company.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Q.* 33, 185–209. doi: 10.2307/3587717

Corder, S. P. (1981). *Error Analysis and Interlanguage*. United States: Oxford University Press.

Costello, E. (2020). 'Requires proofing by a native speaker'–colonization and scholarship. *Insight* 33, 1–7. doi: 10.1629/uksg.502

Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *Ling. Rev.* 27:1. doi: 10.1515/tlir.2010.001

Davies, A. (2003). The native speaker: Myth and realit. *Multilingual Matters*.

Dittmar, N., and Klein, W. (1979). *Developing Grammars: The Acquisition of German Syntax by Foreign Workers: With 9 Figures*. United States: Springer.

Domínguez, L. (2013). *Understanding Interfaces: Second Language Acquisition and First Language Attrition of Spanish Subject Realization and Word Order Variation. Vol. 55*. Amsterdam: John Benjamins Publishing.

Domínguez, L., and Arche, M. J. (2014). Subject inversion in non-native Spanish. *Lingua* 145, 243–265. doi: 10.1016/j.lingua.2014.04.004

Domínguez, L., Hicks, G., and Slabakova, R. (2019). Terminology choice in generative acquisition research: The case of "incomplete acquisition" in heritage language grammars. *Stud. Second. Lang. Acquis.* 41, 241–255. doi: 10.1017/S0272263119000160

Domínguez, L., Tracy-Ventura, N., Arche, M. J., Mitchell, R., and Myles, F. (2013). The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism* 16, 558–577. doi: 10.1017/S1366728912000363

Duffield, N. (2003). Measures of competent gradience. *Lexicon Syntax Int. Second Lang. Acquis.* 30:97. doi: 10.1075/lald.30.06duf

Dulay, H. C., and Burt, M. K. (1974). Natural sequences in child second language acquisition 1. *Lang. Learn.* 24, 37–53. doi: 10.1111/j.1467-1770.1974.tb00234.x

Ellis, R. (1985). Sources of variability in interlanguage. *Appl. Linguis.* 6, 118–131. doi: 10.1093/applin/6.2.118

Ellis, R., Loewen, S., and Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Stud. Second. Lang. Acquis.* 28, 339–368. doi: 10.1Ol7/S027226310606014

Epstein, S. D., Flynn, S., and Martohardjono, G. (1996). Second language acquisition: theoretical and experimental issues in contemporary research. *Behav. Brain Sci.* 19, 677–714. doi: 10.1017/S0140525X00043521

Firth, A., and Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *Mod. Lang. J.* 81, 285–300. doi: 10.1111/j.1540-4781.1997.tb05480.x

Francis, E. J. (2021). *Gradient Acceptability and Linguistic Theory*. Oxford: Oxford University Press.

Friederici, A. D. (2017). *Language in our Brain: The Origins of a Uniquely Human Capacity*. United States: MIT Press.

Gass, S. (1998). Apples and oranges: or, why apples are not orange and don't need to be a response to Firth and Wagner. *Mod. Lang. J.* 82, 83–90. doi: 10.1111/j.1540-4781.1998.tb02597.x

Gibson, E., and Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Lang. Cognit. Proc.* 28, 88–124. doi: 10.1080/01690965.2010.515080

Gregg, K. R. (1996). The logical and developmental problems of second language acquisition. *Handbook Second Lang. Acquis.*, 49–81.

Grüter, T. (2006). Another take on the L2 initial state: evidence from comprehension in L2 German. *Lang. Acquis.* 13, 287–317. doi: 10.1207/s15327817la1304_2

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language 1, 2. *Lang. Learn.* 26, 321–351. doi: 10.1111/j.1467-1770.1976.tb00280.x

Han, Z. (2004). Fossilization in adult second language acquisition. *Multilingual Matters*; March 16, 2004.

Han, Z. (2016). A" reimagined SLA" or an expanded SLA? A rejoinder to The Douglas fir group (2016). *Mod. Lang. J.* 100, 736–740. doi: 10.1111/modl.12378

Han, Z., and Tarone, E. (eds). (2014). *Interlanguage: Forty Years Later. Vol. 39*. Netherlands: John Benjamins Publishing Company.

Hatch, E., and Wagner-Gough, J. (1976). Explaining sequence and variation in second language acquisition. *Lang. Learn.* 4, 39–47.

Heil, J., and López, L. (2020). Acquisition without evidence: English infinitives and poverty of stimulus in adult second language acquisition. *Second. Lang. Res.* 36, 415–443. doi: 10.1177/0267658319850611

Henry, A. (1996). Imperative inversion in Belfast English. *Am. Studies Theory History Ling. Sci. Series* 4, 79–94.

Holliday, A. (2006). Native-speakerism. *ELT J.* 60, 385–387. doi: 10.1093/elt/ccl030

Holliday, A. (2015). "Native-speakerism: taking the concept forward and achieving cultural belief" in *(En) Countering Native-Speakerism.* eds. A. Swan, P. Aboshiha and A. Holliday (London: Palgrave Macmillan), 11–25.

Huebner, T. (1983). Linguistic systems and linguistic change in an interlanguage. *Stud. Second. Lang. Acquis.* 6, 33–53. doi: 10.1017/S0272263100000280

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Lang. Assess. Q.* 8, 229–249. doi: 10.1080/15434303.2011.565844

Hulstijn, J. H., and Andringa, S. (2014). "The effects of age and level of education on the ability of adult native speakers of Dutch to segment speech into words" in *Above and Beyond the Segments*. eds. J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller and E. van Zanten (Netherlands: John Benjamins), 152–164.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Lang. Learn.* 27, 383–410. doi: 10.1111/j.1467-1770.1977.tb00129.x

Hyltenstam, K., and Abrahamsson, N. (2000). Who can become native-like in a second language? All, some, or none? On the maturational constraints controversy in second language acquisition. *Studia Linguistica* 54, 150–166. doi: 10.1111/1467-9582.00056

Jenkins, J., Cogo, A., and Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Lang. Teach.* 44, 281–315. doi: 10.1017/S0261444811000115

Kachru, Y. (1994). Monolingual bias in SLA research. *Tesol Quarterly* 28, 795–800.

Klein, W. (1998). The contribution of second language acquisition research. *Lang. Learn.* 48, 527–549. doi: 10.1111/0023-8333.00057

Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Lang. Var. Chang.* 1, 199–244. doi: 10.1017/S0954394500000168

Kroch, A. (1994). "Morphosyntactic variation" in *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory. Vol. 2*. eds. K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki and E. Zeinfeld (Chicago: Linguistics Society Chicago), 180–201.

Labov, W. (1982). Objectivity and commitment in linguistic science: The case of the black English trial in Ann Arbor. *Lang. Soc.* 11, 165–201. doi: 10.1017/S0047404500009192

Lakshmanan, U., and Selinker, L. (2001). Analysing interlanguage: how do we know what learners know? *Second. Lang. Res.* 17, 393–420. doi: 10.1177/026765830101700406

Lardiere, D. (2003). Revisiting the comparative fallacy: A reply to Lakshmanan and Selinker, 2001. *Second. Lang. Res.* 19, 129–143. doi: 10.1191/0267658303sr216oa

Leal, T., and Slabakova, R. (2019). The relationship between L2 instruction, exposure, and the L2 acquisition of a syntax–discourse property in L2 Spanish. *Lang. Teach. Res.* 23, 237–258. doi: 10.1177/1362168817745714

Leivada, E., and Westergaard, M. (2020). Acceptable ungrammatical sentences, unacceptable grammatical sentences, and the role of the cognitive parser. *Front. Psychol.* 11:364. doi: 10.3389/fpsyg.2020.00364

Leung, C., Harris, R., and Rampton, B. (1997). The idealised native speaker, reified ethnicities, and classroom realities. *TESOL Q.* 31, 543–560. doi: 10.2307/3587837

Lightbown, P., Spada, N., and Wallace, R. (1980). Some effects of instruction on child and adolescent ESL learners. *Res. Second Lang. Acquis.*, 162–172.

Lipsey, M. W. (1990). *Design Sensitivity. Vol. 370*. Newbury Park, CA: Sage.

Long, M. H. (1997). Construct validity in SLA research: A response to Firth and Wagner. *Mod. Lang. J.* 81, 318–323. doi: 10.1111/j.1540-4781.1997.tb05487.x

Lyster, R., and Saito, K. (2010). Interactional feedback as instructional input: A synthesis of classroom SLA research. *Lang. Int. Acquis.* 1, 276–297. doi: 10.1075/lia.1.2.07lys

Mack, M. (1997). *The Monolingual Native Speaker: Not a Norm, but Still a Necessity Studies Linguist. Sci.* 27, 113–146.

Meisel, J. M. (2011). *First and Second Language Acquisition: Parallels and Differences*. United Kingdom: Cambridge University Press.

Meisel, J. M., Clahsen, H., and Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Stud. Second. Lang. Acquis.* 3, 109–135. doi: 10.1017/S0272263100004137

Montrul, S. (2013). The Ups and Downs of the Monolingual Bias and the Bilingual Turn in Second Language Acquisition: Getting Down to Some Specifics. *Talk presented at the Seminar on Second Language Acquisition*. Stockholm University, Sweden; 5 November 2013.

Nemser, W. (1971). Approximative systems of foreign language learners. *IRAL* 9, 115–123. doi: 10.1515/iral.1971.9.2.115

Newmeyer, F. J., and Weinberger, S. H. (1988). "The ontogenesis of the field of second language learning research" in *Linguistic Theory in Second Language Acquisition*. eds. Flynn, S and W. O'Neil (Dordrecht: Springer), 34–45.

Norris, J., and Ortega, L. (2003). "Defining and measuring SLA" in *The Handbook of Second Language Acquisition*. eds. C. J. Doughty and M. C. Long (Texas: Wiley), 716–761.

Ortega, L. (2014). "Ways forward for a bi/multilingual turn in SLA" in *The Multilingual Turn: Implications for SLA, TESOL, and Bilingual Education*. ed. S. May (New York: Routledge/Taylor and Francis), 32–53.

Ortega, L. (2019). SLA and the study of equitable multilingualism. *Mod. Lang. J.* 103, 23–38. doi: 10.1111/modl.12525

Phillips, C. (2009). "Should we impeach armchair linguists?" in *Japanese/Korean Linguistics. Vol. 17*. eds. S. Iwasaki, H. Hoji, P. M. Clancy, and S.-O. Sohn (Stanford, CA: CSLI Publications), 49–64.

Phillipson, R. (1992). ELT: the native speaker's burden? *ELT J.* 46, 12–18. doi: 10.1093/elt/46.1.12

Pica, T. (1983). Methods of morpheme quantification: their effect on the interpretation of second language data. *Stud. Second. Lang. Acquis.* 6, 69–78. doi: 10.1017/S0272263100000309

Pica, T., Kang, H. S., and Sauro, S. (2006). Information gap tasks: their multiple roles and contributions to interaction research methodology. *Stud. Second. Lang. Acquis.* 28, 301–338. doi: 10.1017/S027226310606013X

Polinsky, M., and Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Lang. Ling. Compass* 1, 368–395. doi: 10.1111/j.1749-818X.2007.00022.x

Quené, H. (2010). How to design and analyze language acquisition studies. *Exp. Methods Lang. Acquis. Res.* 27:269. doi: 10.1075/lllt.27.15que

Rankin, T., and Whong, M. (2020). *Grammar: A Linguists' Guide for Language Teachers*. England: Cambridge University Press.

Schachter, J. (1986). In search of systematicity in interlanguage production. *Stud. Second. Lang. Acquis.* 8, 119–133. doi: 10.1017/S0272263100006045

Schachter, J. (1988). Second language acquisition and its relationship to universal grammar. *Appl. Linguis.* 9, 219–235. doi: 10.1093/applin/9.3.219

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Int. Studies Bulletin* 6, 93–124. doi: 10.1177/026765839000600201

Schütze, C. T. (1996/2016). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Berlin: Language Science Press.

Schwartz, B. D. (1997). On the basis of the basic variety. *Second. Lang. Res.* 13, 386–402. doi: 10.1191/026765897670070857

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *Int. J. Appl. Linguist.* 11, 133–158. doi: 10.1111/1473-4192.00011

Selinker, L. (1972). Interlanguage. *IRAL* 10, 209–231.
Slabakova, R., Rothman, J., and Kempchinsky, P. (2011). Gradient competence at the syntax-discourse interface. *EUROSLA Yearb.* 11, 218–243. doi: 10.1075/eurosla.11.12rot
Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in non native grammars of Italian. *Second. Lang. Res.* 9, 22–47. doi: 10.1177/026765839300900102
Sorace, A. (1996). "The use of acceptability judgments in L2 acquisition research," in *The Handbook of Second Language Acquisition*. eds. W. Ritchie and T. Bhatia (New York: Academic Press), 375–409.
Sorace, A., and Keller, F. (2005). Gradience in linguistic data. *Lingua* 115, 1497–1524. doi: 10.1016/j.lingua.2004.07.002
Sprouse, J. (2011). A test of the cognitive assumptions of Magnitude Estimation: Commutativity does not hold for acceptability judgments. *Language.* 87, 274–288.
Sprouse, J., Schütze, C. T., and Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from linguistic inquiry 2001–2010. *Lingua* 134, 219–248. doi: 10.1016/j.lingua.2013.07.002
Stauble, A. M. E. (1981). *A Comparative Study of a Spanish-English and Japanese-English Second Language Continuum: Verb Phrase Morphology*. Los Angeles: University of California.
Tarone, E. (1979). Interlanguage as chameleon 1. *Lang. Learn.* 29, 181–191. doi: 10.1111/j.1467-1770.1979.tb01058.x
Tarone, E. (2018). "Interlanguage" in *The Encyclopedia of Applied Linguistics*. ed. C. A. Chapelle (Texas: Wiley), 1–7.
Tarone, E., Frauenfelder, U. H., and Selinker, L. (1976). Systematicity/variability and stability/instability in interlanguage systems. *Lang. Learn.* 4, 93–134.
The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *Mod. Lang. J.* 100(S1), 19–47. doi:10.1111/modl.12301.
White, L. (2003). *Second Language Acquisition and Universal Grammar*. England: Cambridge University Press.
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
Whong, M., Gil, K.-H., and Marsden, H. (2013). *Universal Grammar and the Second Language Classroom*. Dordrecht, Netherlands: Springer.
Yang, C. D. (2002). *Knowledge and Learning in Natural Language*. England: Oxford University Press on Demand.
Zuengler, J., and Cole, K. (2005). Language socialization and second language learning. *Handbook Res. Second Lang. Teach. Learn.* 1, 301–316. doi: 10.4324/9781410612700
Zuengler, J., and Miller, E. R. (2006). Cognitive and sociocultural perspectives: two parallel SLA worlds? *TESOL Q.* 40, 35–58. doi: 10.2307/40264510

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 10 November 2021
doi: 10.3389/fpsyg.2021.742127

Going Native? Yes, If Allowed by Cross-Linguistic Similarity

Gillen Martínez de la Hidalga, Adam Zawiszewski and Itziar Laka*

Department of Linguistics and Basque Studies, Faculty of Arts, University of the Basque Country, Vitoria-Gasteiz, Spain

Can native competence be achieved in a second language? Here, we focus on the Language Distance Hypothesis that claims that early and proficient bilinguals can achieve native competence for grammatical properties shared by their two languages, whereas unshared grammatical properties pose a challenge for native-like syntactic processing. We present a novel behavioral and Event-Related Potential (ERP) study where early and proficient bilinguals behave native-like in their second language when processing (a) argument structure alternations in intransitive sentences involving agent vs. patient subjects and (b) subject verb agreement, both of which are grammatical properties shared by their two languages of these bilinguals. Compared to native Basque bilinguals (L2Spanish) on the same tasks, non-natives elicited similar sentence processing measures: (a) in the acceptability task they reacted faster and more accurately to unaccusative sentences than to unergatives and to person than number violations: (b) they generated a larger P600 for agreement violations in unaccusative sentences than unergatives; (c) they generated larger negativity and positivity effects for person than for number violations. Previous studies on Basque-Spanish bilinguals find that early and proficient non-natives display effects distinct from natives in both languages when processing grammatical properties where Basque and Spanish diverge, such as argument alignment (ergative/nominative) or word order type (OV/VO), but they perform native-like for shared properties such as subject agreement and word meaning. We contend that language distance, that is, the degree of similarity of the languages of the bilingual is a crucial factor that deserves further and detailed attention to advance our understanding of when and how bilinguals can *go native* in a second language.

Keywords: non-native language processing, event-relate potentials, unergative vs. unaccusative predicates, subject-verb agreement, phi-features, bilingualism

OPEN ACCESS

Edited by:
Pedro Guijarro-Fuentes, University of the Balearic Islands, Spain

Reviewed by:
Nuria Sagarra, Rutgers, The State University of New Jersey, United States
Maria Garraffa, University of East Anglia, United Kingdom

****Correspondence:***
Adam Zawiszewski
adam.zawiszewski@ehu.eus

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

Received: *15 July 2021*
Accepted: *07 October 2021*
Published: *10 November 2021*

Citation:
Martínez de la Hidalga G, Zawiszewski A and Laka I (2021) Going Native? Yes, If Allowed by Cross-Linguistic Similarity. Front. Psychol. 12:742127. doi: 10.3389/fpsyg.2021.742127

INTRODUCTION

Can non-native speakers attain native-like competence in grammatical processing? Research carried out throughout the last decades has identified key factors to take into account when studying non-native syntactic processing, namely age of acquisition (AoA), proficiency, similarity between L1 and L2 and active use of the language (Caffarra et al., 2015; Hartshorne et al., 2018; Brice et al., 2019).

In second language acquisition, syntax is reported to be harder to acquire than other aspects of language (Ojima et al., 2005; Kotz, 2009; Vandenberghe et al., 2019). It has also been shown that AoA and the level of proficiency play a big role in attaining native-like performance

(i.e., Weber-Fox and Neville, 1996; Wartenburger et al., 2003). Weber-Fox and Neville (1996), for instance, used ERPs to test Chinese-English adult bilinguals exposed to English at different age during the life span (1–3, 4–6, 7–10, 11–13, and after 16 years of age) and asked the participants to read sentences containing syntactic and semantic anomalies. Results revealed significant AoA effects for syntactic processing (phrase structure, specificity and subjacency constraints), that is, in comparison to English monolinguals, behavioral and electrophysiological measures of Chinese-English bilinguals were affected by a delay in L2 exposure as short as 1–3 years. By contrast, regarding semantic anomalies, only subjects exposed to English after 11–13 years showed differences as compared to natives. Wartenburger et al. (2003), in turn, used the fMRI method to test the effects of AoA and proficiency in three groups of Italian-German bilinguals who learned their L2 at different ages and had different proficiency levels (early AoA (= at birth), high proficiency group; late AoA (>6 years), high proficiency group and late AoA (>6 years), low proficiency group). Participants read sentences in their L1 and L2 containing syntactic (gender, number or case disagreement) and semantic anomalies (i.e., "The deer shoots the hunter"). Results revealed that differences in the fMRI pattern reported for the syntactic task were due to the delay in AoA, while the pattern of brain activity for semantic judgment depended on the level of proficiency, supporting the findings of Weber-Fox and Neville (1996).

On the other hand, several studies affirm that high proficiency L2 speakers can attain native-like performance (Friederici et al., 2002; Rossi et al., 2006; Hernandez et al., 2007; Kotz et al., 2008) regardless of their late AoA, thus challenging the Critical Period Hypothesis (Lenneberg, 1967). Friederici et al. (2002) used ERP measures to test adult German learners (AoA = 24.1 years) of an artificial language and showed a similar ERP pattern to that reported for native speakers of German on a similar task in natural language. According to the authors, these results indicate that a language learned late can be processed in a native-like way. Similarly, Rossi et al. (2006) showed that high-proficiency late L2 Italian-German and German-Italian learners (AoA > 10 years) display the same ERP components as native speakers when processing word category and subject-verb agreement syntactic violations, suggesting that with a high proficiency L2 learners can show native-like responses regardless of the late AoA.

Finally, some studies report differences between native and non-natives regarding certain syntactic phenomena but not others (Zawiszewski et al., 2011; Foucart and Frenck-Mestre, 2012; Erdocia et al., 2014; Díaz et al., 2016; Zawiszewski and Laka, 2020). Zawiszewski et al. (2011), Díaz et al. (2016), and Zawiszewski and Laka (2020) examined the processing of case morphology and Erdocia et al. (2014) the processing of word order comparing native and non-native speakers of Basque (L1Spanish) and found that non-natives, despite an early AoA and high competence in their L2 did not process these two aspects of Basque grammar like natives. Since case alignment and basic word order are two main grammatical features that Basque and Spanish do not share (Basque is ergative and OV, Spanish is nominative and VO), they concluded that linguistic distance, that is, the degree of similarity of the bilingual's grammars was a relevant factor in final attainment in second language processing.

These different sets of findings reported in the L2 processing literature have been accounted by many theoretical proposals. Some posit that L2 acquisition strongly depends on the L1 and thus the results can be interpreted in terms of a positive or negative transfer (i.e., The Unified Model of Language Acquisition, Hernandez et al., 2005; MacWhinney, 2005). Schwartz and Sprouse (1996) also suggest that L2 acquisition hinges on the features available in the L1 (the Full Transfer/Full Access model) and this view is also compatible with the Interpretability Hypothesis (Tsimpli and Dimitrakopoulou, 2007), which posits that only interpretable features are accessible to the L2 learners while the uninterpretable ones are subject to critical period constraints and, consequently, inaccessible to L2 learners. The possibility of syntactic information being shared between both languages has been also suggested by Hartsuiker et al. (2004) (Shared Syntax Account): it suggests that grammatical rules that are the same in L1 and L2 are represented once. In other words, L2 learners would rely on their L1 whenever using a grammatical structure present in the two languages (see also Zawiszewski and Laka, 2020, for similar assumptions), Conversely, Clahsen and Felser (2006) put forward the Shallow Structure Hypothesis and suggest that late L2 learners are not able to process syntax in a native-like way and have to rely to a large extent on semantic/pragmatic information (see also Steinhauer et al., 2009 for a discussion).

THE PRESENT STUDY

The present study sought to examine the processing of intransitive predicates in Basque by early and proficient L1Spanish—L2Basque bilinguals. To this purpose, we used grammatical and ungrammatical person and number agreement manipulations and compared the results to those previously reported by Martinez de la Hidalga et al. (2019) for natives.

The distinction between intransitives whose sole argument is an agent (unergatives) and intransitives whose sole argument is a theme (unaccusatives) is a general property of grammars [Unaccusative Hypothesis (UH), Perlmutter, 1978], and both Basque and Spanish differentiate these two types of predicates. More precisely, the UH claims that unaccusative involve more complex derivations than unergatives because themes are promoted to subjects or undergo movement and leave a trace (Burzio, 1986), whereas agents are born as subjects. Importantly, some authors propose for the unaccusative verbs in Basque the same derivation as stated by the UH (Ortiz de Urbina, 1989), while others claim no need for the extra derivational step (Laka, 2006a,b; Levin, 1983). More complex derivations are usually related to a greater processing cost (longer reading or reaction times, larger ERP signatures) as compared to less complex structures (i.e., Matzke et al., 2002). Consequently, larger processing cost is expected for unaccusatives in comparison to unergatives. Subject agreement is also a shared property between Basque and Spanish, and both grammars represent it by means of person and number features. Unaccusative verbs have been

found to be harder to learn than unergatives for second language learners at initial stages (Yuan, 1999; Oshita, 2001; Montrul, 2005; *inter alia*). Oshita (2001) for instance, put forth the Unaccusative Trap Hypothesis (UTH), arguing that L2 learners assume at first all intransitive predicates to be unergatives. As proficiency increases, however, learners notice that unaccusatives function differently and start making differences between the two predicates, and at higher levels of proficiency they are found to perform native-like regarding this linguistic dimension.

In a previous study carried out in Basque, Martinez de la Hidalga et al. (2019) investigated Basque-Spanish bilinguals in order to test the UH hypothesis and phi-feature processing. Results revealed that in the acceptability task the participant reacted faster and more accurately to unaccusative sentences than to unergatives and to person than number violations and they generated a larger P600 for agreement violations in unaccusative sentences than in unergatives. Furthermore, they generated larger negativity and positivity effects for person than for number violations. Overall, the results revealed greater processing costs for unergatives than for unaccusatives and the authors interpreted these findings as evidence providing support for different structural representations of both types of predicates. However, the prediction of higher processing cost for unaccusatives than for unergatives was not confirmed, supporting the idea of an inherent rather than structural nature of case in Basque (Levin, 1983; Laka, 2006a,b). Regarding agreement features, native speakers processed person and number features separately, the person being far more salient than the number (see Carminati, 2005; Zawiszewski et al., 2016; Mancini, 2018, for more information on the processing of person and number features).

Hypotheses and Predictions

Our working hypothesis is the Language Distance Hypothesis (LDH, after Zawiszewski and Laka, 2020): no differences are expected for processing traits of L2 that are present in L1, whereas even at an early AoA and high proficiency in L2, native vs. non-native differences will arise in the processing grammatical properties of L2 not present in L1. Previous studies in Basque investigating ergative case morphology (Díaz et al., 2011; Zawiszewski et al., 2011; Zawiszewski and Laka, 2020) and word order processing (Erdocia et al., 2014) in native and early and highly proficient Spanish-Basque bilinguals found differences between both populations, attributed by the authors to the diverging grammatical characteristics of Basque and Spanish.

In the present study, the experimental manipulations involve grammatical traits shared by both Basque and Spanish, namely the distinction between unaccusative vs. unergative predicates, and person vs. number features in subject-verb agreement. However, despite the fact that both Basque and Spanish distinguish between unaccusative and unergative predicates, in Basque agents bear an ergative case marking and themes are morphologically unmarked. In contrast, in Spanish all subjects are morphologically indistinguishable.

We tentatively hypothesize that, given the early AoA and high proficiency of the non-native speakers under study, a similar pattern of results to that reported in Martinez de la Hidalga et al. (2019) will emerge: (a) faster and more accurate responses to unaccusative sentences than to unergative ones and to person violations than number violations in the acceptability task; (b) a general N400–P600 pattern as an ERP response to verb agreement violations, unaccusative violations generated a larger positivity as compared to unergatives and person feature violations generated a larger negativity as compared to number feature violations in the early time window; and (c) person violations in the unergative condition generated a larger positivity as compared to number violations, and larger positivity obtained for number violations in the unaccusative condition as compared to number violations in the unergative condition in the late time window (P600 effect). The predictions made by the LDH are also compatible with the Shared Syntax account (Hartsuiker et al., 2004).

Participants

In this experiment 26 early and highly proficient non-native speakers of Basque, whose L1 was Spanish[1] took part in the experiment (five males; mean age 20.5 years, *SD* = 2.67; AoA = 3.31 years, *SD* = 1.3). Data from two participants were excluded as a result of excessive eye movements and other artifacts. All participants were schooled in Basque from early childhood and were therefore highly proficient in Basque (see **Table 1** for details) as revealed by the fact that 21 participants had a certified C1 level in Basque and the remaining 3 were completing their undergraduate degree in Basque.

MATERIALS AND METHODS

416 sentences distributed in four lists (256 experimental and 160 fillers) were created. The materials were organized according to the manipulations used in the experiment (2 × 2 × 2 design): predicate type (unaccusative vs. unergative), feature (person and number), and grammaticality (grammatical and ungrammatical) (see **Table 2**). For person conditions 2nd person was used in the grammatical condition and for 1st person was used in the ungrammatical manipulation. For number conditions, the design used in Mancini et al. (2011) was followed: 3rd singular vs. plural manipulations. The critical words were the auxiliary verbs, always preceded by the main verbs and followed by three words all verbs were controlled for length and frequency.

Procedure

Personal computers (Windows 7 operating system) and Presentation software (version 16.3) were used to present the stimuli on screen. Before the experiment started, participants were told about the EEG procedure and seated comfortably in a quiet room in front of a 24 inch monitor. The experiment was conducted in the Experimental Linguistics Laboratory at the University of the Basque Country (UPV/EHU) in Vitoria-Gasteiz. Participants conducted an acceptability judgment task, were both accuracy and reaction times were recorded. Sentences were displayed in the middle of the screen word by word for 350 ms (ISI = 250). A fixation cross (+) indicated the

[1]One participant had Catalan as mother tongue; given the typological similarity between Catalan and Spanish, those data were not discarded from the final analysis.

TABLE 1 | The following seven-point scale was applied for measuring the relative use of language: 1 = I speak only Basque, 2 = I speak mostly Basque, 3 = I speak Basque 75% of the time, 4 = I speak Basque and Spanish with similar frequency, 5 = I speak Spanish 75% of the time, 6 = I speak mostly Spanish, 7 = only Spanish.

Relative use of language		
Before primary school (0–3 years)		6.54 (0.72)
Primary school (4–12 years)		
School		2.88 (1.42)
Home		6.54 (0.72)
Others		5.86 (1.08)
Secondary school (12–18 years)		
School		3.58 (1.07)
Home		6.67 (0.56)
Others		5.92 (0.93)
At time of testing		
University/work		4.5 (1.69)
Home		6.63 (0.58)
Others		5.46 (1.02)
Self-rated proficiency	**Basque**	**Spanish**
Speaking	5.92 (0.58)	6.88 (0.34)
Comprehension	6.42 (0.5)	6.92 (0.28)
Reading	6.46 (0.51)	6.83 (0.38)
Writing	6 (0.59)	6.67 (0.48)

Proficiency level was determined by using the following four-point scale: 7 = native-like proficiency,6 = full proficiency, 5 = working proficiency, 4 = limited proficiency. SDs values are in parentheses.

beginning of each sentence trial. After each trial the words *zuzen?* "correct?" or *over?* "incorrect?" were displayed in the screen, and participants had to judge the acceptability of the previously shown sentence as either correct or incorrect. Half of participants used the left hand for correct responses (left Ctrl) and the other half the right hand (right Intro).

All sentences were randomly distributed in four blocks. Each block lasted approximately 10 min each and participants had a short break between each block, for as long as they needed. Before the experiment began, participants ran a short training session consisting of three trials. They were instructed to avoid blinking or moving while the sentences were being displayed and to make the acceptability judgment as fast and accurately as possible. The whole experiment, including electrode-cap application and removal, lasted about 1 h 15 m.

EEG Recording

The EEG was recorded from 32 active electrodes secured in an elastic cap (Acticap System, Brain Products). Electrodes were set on standard positions according to the extended Internationals 10–20 system accordingly: Fp1/Fp2, Fz, F3/F4, F7/F8, FC5/FC6, FC1/FC2, T7/T8, C3/C4, Cz, CP5/CP6, CP1/CP2, P7/P8, P3/P4, Pz, O1/02, Oz, LM, VEOG and HEOG. All recordings were referenced to right mastoid position and re-referenced off-line to the linked mastoids. Vertical and horizontal eye movements and blinks were monitored by means of two electrodes positioned beneath and to the right of the right eye. Electrode impedance was kept below 5 kOhm at all scalp and below 10 kOhm for the eye electrodes. The electrical signals were digitized online at a rate of 500 Hz by a Brain Vision amplifier system and filtered offline within a band pass of 0.1–35 Hz. After the EEG data were recorded, the ocular correction procedure (Gratton et al., 1983) as well as the artifact rejection procedure were applied (offline). Trials with other artifacts with any voltage exceeding 150 μV and voltage steps between two sampling points exceeding 35 μV were removed.

TABLE 2 | Experimental conditions with examples of experimental materials.

Conditions			**Sentence examples**
Predicate type	**Feature**	**Grammaticality**	
Unaccusative	Person	Grammatical	1. Zu gaur goizean bueltatu zara Bilbotik. you-ABS today morning.in returned 2SG.ABS-be Bilbao-from "You have come back from Bilbao this morning."
		Ungrammatical	2. *Zu gaur goizean bueltatu naiz Bilbotik. you-ABS today morning.in returned 1SG.ABS-be Bilbao-from
	Number	Grammatical	3. Hura gaur goizean bueltatu da Bilbotik. 3.SG-ABS today morning.in returned 3SG.ABS-be Bilbao-from
		Ungrammatical	4. *Hura gaur goizean bueltatu dira Bilbotik. 3.SG-ABS today morning.in returned 3PL.ABS-be Bilbao-from
Unergative	Person	Grammatical	5. Zuk goizean biziki sufritu duzu aurkezpenean. you-ERG morning a.lot suffered have-2SG.ERG presentation-the-at "You have suffered a lot this morning at the presentation."
		Ungrammatical	6. *Zuk goizean biziki sufritu dut aurkezpenean. you-ERG morning a.lot suffered have-1SG.ERG presentation-the-at
	Number	Grammatical	7. Hark goizean biziki sufritu du aurkezpenean. 3.SG-ERG morning a.lot suffered have-3SG.ERG presentation-the-at
		Ungrammatical	8. *Hark goizean biziki sufritu dute aurkezpenean. 3.SG-ERG morning a.lot suffered have-3PL.ERG presentation-the-at

**stands for ungrammatical sentences.*

Data Analysis

For the data analysis four types of subject agreement violations were compared: unaccusative person violations (*zara* "be.2SG" vs. **naiz* "be.1SG"; conditions 1 vs. 2 in **Table 1**, respectively); unaccusative number violations (*da* "be.3SG" vs. **dira* "be.3PL"; conditions 3 vs. 4 in **Table 1**, respectively); unergative person violations (*duzu* "have.2SG" vs. **dut* "have.1SG"; conditions 5 vs. 6 in **Table 1**, respectively); unergative number violations (*du* "have.3SG" vs. **dute* "have.3PL"; conditions 7 vs. 8 in **Table 1**, respectively).

For the ERP measures, segments were created from 200 ms before and 1,000 ms after the onset of the critical words (the auxiliary) in the sentences. The trials associated with each sentence type were averaged for each participant. The EEG 200 ms prior to the onset was also used as a baseline for all sentence type comparisons.

Three hundred to four hundred milliseconds and four hundred to seven hundred milliseconds temporal windows were selected for statistical analysis in all conditions based on the literature and visual inspection of the data. After the stimuli were recorded and averaged, analyses of variance (ANOVA) were carried out in nine regions of interest that were computed out of 27 electrodes: lateral electrodes: left frontal (F7, F3, FC5), left central (T7, FP5, C3), left parietal (P7, P3, O1), right frontal (F4, F8, FC6), right central (C4, FP6, T8), and right parietal (P8, P4, O2); midline electrodes: frontal (Fp1, Fz, Fp2), central (FC1, Cz, FC2), and parietal (CP1, Pz, CP2). Repeated-measures ANOVAs were conducted in all experimental manipulations and trials (correctly and incorrectly judged trials) for each window of time using five within-subjects factors: grammaticality (2 levels: grammatical, ungrammatical), type (2 levels: unaccusative, unergative), feature (2 levels: person, number), hemisphere (2 levels: left, right), and region (3 levels: frontal, central and parietal). Midline (frontal, central, and parietal) electrodes were analyzed independently. Whenever the sphericity of variance was violated (Greenhouse and Geisser, 1959) correction was applied to all the data with greater than one degree of freedom in the numerator. Finally, further statistical comparisons were carried out (split by the grammaticality condition) whenever we found a statistically significant interaction. We only consider effects for the type, feature, hemisphere or region factors when there is an interaction with grammaticality.

For the behavioral results, error rates and response latencies of all the trials repeated measures ANOVAs were performed with grammaticality (two levels: grammatical, ungrammatical), type (two levels: unaccusative, unergative) and feature (two levels: person, number) conditions as within-subject factors. Subsequent comparisons (by subject and by item) were carried out whenever a grammatical interaction was significant.

RESULTS

Behavioral Results

Here, results concerning the acceptability task and reaction times are presented. Participants were very accurate in the acceptability task (mean accuracy of 91.84%, SDE = 1.3), as was to be expected given their high proficiency in Basque (see **Figure 1**).

Regarding acceptability judgment errors, the analysis showed a marginally significant GRAMMATICALITY effect in the analysis by item [$F1(1, 23) = 1.8$, $p = 0.193$; $F2(1, 253) = 3.03$, $p = 0.083$] revealing higher accuracy for the ungrammatical sentences as compared to the grammatical ones (92.74% vs. 90.95%). The analysis of accuracy also revealed a main FEATURE effect [$F1(1, 23) = 23.4$, $p < 0.001$; $F2(1, 253) = 24.62$, $p < 0.001$] indicating that participants were more accurate with conditions containing person feature (94.04%) compared to conditions containing number feature (89.65%).

The GRAMMATICALITY*FEATURE interaction turned out to be statistically significant as well [$F1(1, 23) = 5.34$, $p = 0.03$; $F2(1, 253) = 3.22$, $p = 0.074$]. The analyses by grammaticality factor showed that participants were significantly less accurate with grammatical person (92.45%) than with ungrammatical person (95.63%) [$F1(1, 23) = 5.97$, $p = 0.023$; $F2(1, 253) = 7.63$, $p = 0.006$], whereas there were no differences between grammatical number (89.46%) and ungrammatical number (89.85%) [$F1(1, 23) = 0.06$, $p = 0.81$; $F2(1, 253) = 0.01$, $p = 913$]. The analyses by feature factor showed that participants were more accurate with grammatical person (92.45%) than with grammatical number (89.46%) [$F1(1, 23) = 6.31$, $p = 0.02$; $F2(1, 253) = 5.74$, $p = 0.017$], and they were significantly more accurate with ungrammatical person (95.63%) than with ungrammatical number (89.85%) [$F1(1, 23) = 35.2$, $p < 0.001$; $F2(1, 253) = 23.56$, $p < 0.001$].

Finally, a triple TYPE*GRAMMATICALITY*FEATURE was significant in the analysis by subject [$F1(1, 23) = 8.09$, $p = 0.009$; $F2(1, 253) = 2.77$, $p = 0.097$]. The analyses by grammaticality factor showed that in unaccusatives grammatical person condition (94.54%) did not differ from ungrammatical person condition (95.16%) [$F(1, 23) = 0.19$, $p = 0.667$], and neither did grammatical and ungrammatical number (89.98% vs. 90.37%) [$F(1, 23) = 0.07$, $p = 0.788$]. In the unergative conditions participants were significantly more accurate with sentences containing ungrammatical person (96.1%) than with grammatical person (90.37%) [$F(1, 23) = 13.32$, $p = 0.001$], but no differences were found between grammatical (88.93%) and ungrammatical number (89.32%) [$F(1, 23) = 0.03$, $p = 0.861$]. The analyses by type factor revealed participants were more accurate with sentences containing grammatical person feature in unaccusatives (94.54%) than in unergatives (90.37%) [$F(1, 23) = 12.38$, $p = 0.002$], whereas no differences were found between sentences containing ungrammatical person feature in unaccusatives (95.16%) and in unergatives (96.1%) [$F(1, 23) = 0.66$, $p = 0.423$]. With regard to number feature, no differences were found between grammatical unaccusative (89.98%) and (88.93%) unergative predicates, and neither between ungrammatical unaccusative (90.37%) and unergative (89.33%) predicates. Finally, the analyses by feature factor showed that participants were significantly more accurate with grammatical unaccusative sentences containing person feature (94.54%) than with number feature (89.98%) [$F(1, 23) = 15.89$, $p = 0.001$], and similarly ungrammatical unaccusative sentences containing person feature (95.16%) were judged more accurately

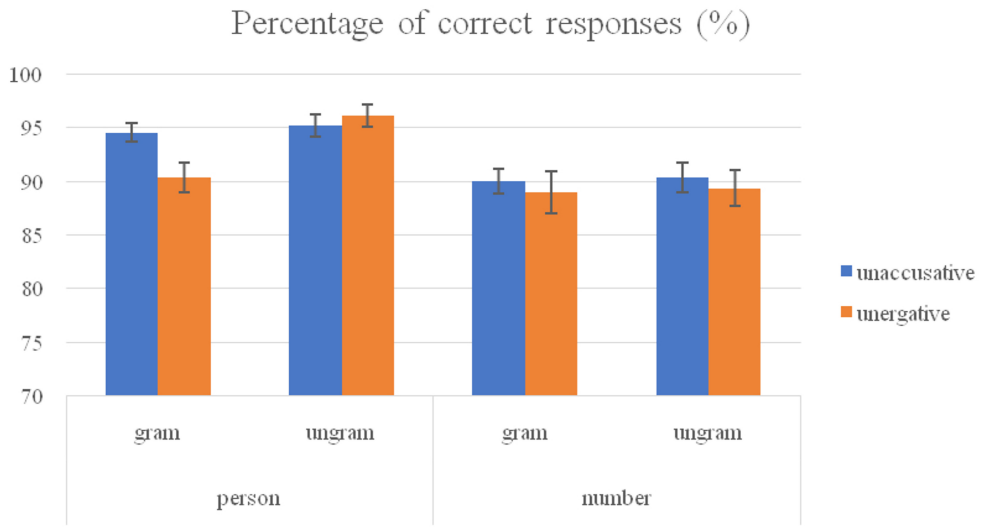

FIGURE 1 | Percentage of correct responses (%) and standard deviation error (SDE) of non-native speakers of Basque.

than number feature (90.37%) [$F(1, 23) = 13.65$, $p = 0.001$]. Regarding unergative predicates, no differences were found between grammatical sentences containing person and number feature [$F(1, 23) = 0.74$, $p = 0.397$], but ungrammatical sentences containing person feature (96.1%) were judged significantly more accurately than ungrammatical sentences containing number feature (89.33%) [$F(1, 23) = 29.67$, $p < 0.001$].

Regarding response times (see **Figure 2**), the analyses revealed a main TYPE effect [$F1(1, 23) = 16.41$, $p = 0.001$; $F2(1, 253) = 4.21$, $p = 0.041$] indicating participants reacted faster to unaccusative predicates (668.37 ms) than to unergative predicates (707.25 ms). A main GRAMMATICALITY effect [$F1(1, 23) = 71.51$, $p < 0.001$; $F2(1, 253) = 207.2$, $p < 0.001$] revealed that participants were significantly faster reacting to ungrammatical sentences (597.93 ms) compared to their grammatical counterparts (777.68 ms). A FEATURE effect [$F1(1, 23) = 11.16$, $p = 0.003$; $F2(1, 253) = 9.61$, $p = 0.002$] revealed that participants were significantly faster responding to sentences containing person feature (665.47 ms) than number feature (710.15 ms).

ERP Results

After the baseline correction, epochs with artifacts were rejected, which resulted in the exclusion of approximately 6.91% ($SD = 2.43$) of the trials. Similarly to the procedure reported in Martinez de la Hidalga et al. (2019), 300–400 ms. time window was selected for an early time window and a 400–700 ms. time window was chosen as a late time window.

Regarding the early time window (300–400 ms), the analysis of the lateral electrodes revealed a main GRAMMATICALITY effect [$F(1, 23) = 18.92$, $p < 0.001$] indicating a larger negativity for the ungrammatical conditions as compared to the grammatical ones (1.08 μV vs. 2 μV).

Regarding the midline electrodes, a main effect of GRAMMATICALITY showed that overall ungrammatical conditions (2.04 μV) displayed a larger negativity than grammatical conditions (2.93 μV) [$F(1, 23) = 11.13$, $p = 0.003$]. A significant TYPE*GRAMMATICALITY interaction was found [$F(1, 23) = 4.9$, $p = 0.037$]. Further analysis (by grammaticality) showed no significant differences between ungrammatical (2.36 μV) and grammatical unaccusatives (2.79 μV) [$F(1, 23) = 2.25$, $p = 0.147$] but a larger negativity for the ungrammatical unergative condition (1.73 μV) in comparison to the grammatical unergative condition (3.07 μV) [$F(1, 23) = 12.67$, $p = 0.002$] was found. The comparison by type revealed no differences between the grammatical unaccusative (2.79 μV) and unergative (3.07 μV) conditions [$F(1, 23) = 0.76$, $p = 0.394$], and neither between ungrammatical unaccusative (2.36 μV) and unergative (1.73 μV) conditions [$F(1, 23) = 1.66$, $p = 0.211$].

The analysis of the lateral electrodes in the late time window (400–700 ms) revealed a main GRAMMATICALITY effect [$F(1, 23) = 60.25$, $p < 0.001$] indicating a larger positivity for the ungrammatical conditions as compared to the grammatical ones (2.08 μV vs. −0.03 μV). In addition, a significant main effect of FEATURE emerged [$F(1, 23) = 13.47$, $p = 0.001$], indicating that overall person feature generated a larger positivity as compared to number feature (1.44 μV vs. 0.61 μV).

A significant TYPE*GRAMMATICALITY interaction was found [$F(1, 23) = 9.34$, $p = 0.006$]. Further analysis (by grammaticality) showed a significantly larger positivity for the ungrammatical unaccusative condition (2.32 μV) in comparison to the grammatical one (−0.18 μV) [$F(1, 23) = 64.23$, $p < 0.001$] and a larger positivity for the ungrammatical unergative condition (1.83 μV) in comparison to the grammatical unergative condition (0.12 μV) [$F(1, 23) = 35.3$, $p < 0.001$]. The comparison by type revealed no differences between the grammatical unaccusative (−0.18 μV) and unergative (0.12 μV) conditions [$F(1, 23) = 1.21$, $p = 0.282$] and no differences emerged for ungrammatical unaccusative manipulations (2.32 μV) in

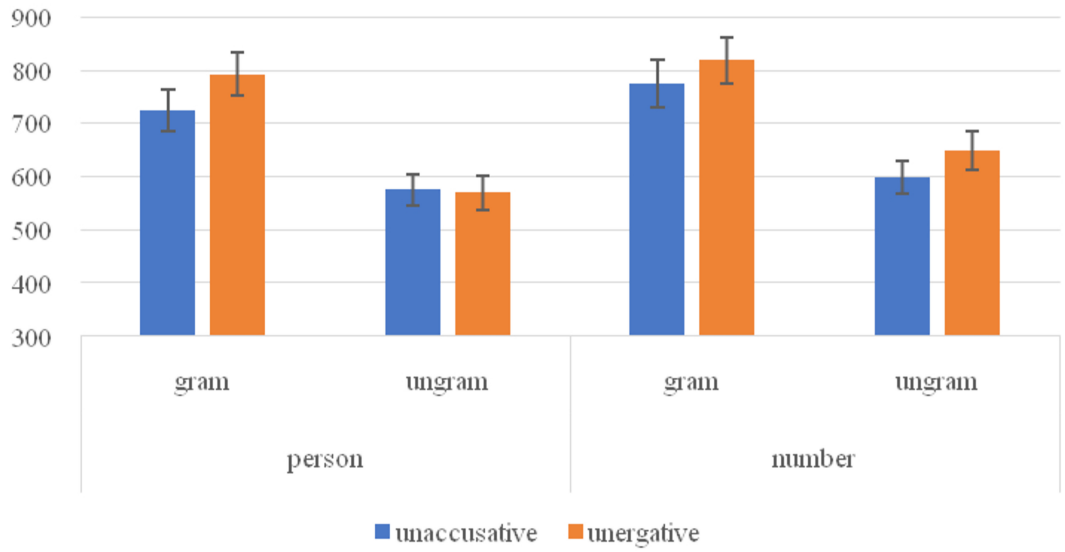

FIGURE 2 | Mean reaction times (ms) and standard deviation error (SDE) of non-native speakers of Basque.

comparison to the unergative manipulations (1.83 μV) [$F(1, 23) = 2.28$, $p = 0.145$].

Regarding the midline electrodes, a main effect of GRAMMATICALITY showed that overall ungrammatical conditions (3.59 μV) displayed a larger positivity than grammatical conditions (0.63 μV) [$F(1, 23) = 59.63$, $p < 0.001$]. In addition, a significant FEATURE effect emerged [$F(1, 23) = 20.94$, $p < 0.001$], indicating that overall person feature generated a larger positivity as compared to number feature (2.78 μV vs. 1.44 μV).

A significant TYPE*GRAMMATICALITY interaction was found [$F(1, 23) = 13.45$, $p = 0.001$]. Further analysis (by grammaticality) showed a significantly larger positivity for the ungrammatical unaccusative condition (4.03 μV) in comparison to the grammatical one (0.36 μV) ($F(1, 23) = 56.96$, $p < 0.001$) and a larger positivity for the ungrammatical unergative condition (3.14 μV) in comparison to the grammatical number condition (0.9 μV) [$F(1, 23) = 37.99$, $p < 0.001$]. The comparison by type revealed no differences between grammatical unergatives (0.9 μV) and unaccusatives (0.36 μV) [$F(1, 23) = 3.02$, $p = 0.096$], but a slightly larger positivity emerged for ungrammatical unaccusative manipulations (4.03 μV) in comparison to the unergative manipulations (3.14 μV) [$F(1, 23) = 3.91$, $p = 0.06$]. See **Figure 3** for the grand average patterns, **Figure 4** for the mean voltage difference maps and **Table 3** for the summary of the results.

Native and Non-native Comparison

In order to better understand the similarities and differences between the non-natives and the native speakers tested in Martinez de la Hidalga et al. (2019), we performed an additional analysis comparing both groups directly.

Behavioral Results

Regarding accuracy, no differences between both groups were found. A marginal main effect of TYPE emerged [$F1(1, 46) = 2.85$, $p = 0.098$; $F2(1, 252) = 4.18$, $p = 0.041$] indicating that overall, both native and non-native participants were more accurate with conditions containing unaccusative predicates (92.76%) compared to unergative predicates (91.82%). The analysis of accuracy revealed a significant main GRAMMATICALITY effect [$F1(1, 46) = 4.89$, $p = 0.032$; $F2(1, 252) = 13.49$, $p < 0.001$] revealing that overall both native and non-native participants were more accurate with conditions containing ungrammatical sentences (93.41%) compared to grammatical sentences (91.16%). The analysis of accuracy also revealed a main FEATURE effect [$F(1, 46) = 41.51$, $p < 0.001$; $F2(1, 252) = 41.5$, $p < 0.001$] suggesting that both natives and non-natives were more accurate with conditions containing person feature (94.17%) compared to conditions containing number feature (90.4%).

A GRAMMATICALITY*FEATURE interaction turned out to be marginally significant in the by subject analysis [$F1(1, 23) = 3.82$, $p = 0.057$; $F2(1, 252) = 2.4$, $p = 0.118$]. The analyses by grammaticality factor showed that participants were significantly less accurate with grammatical person (92.61%) than with ungrammatical person (95.73%) [$F(1, 46) = 11.57$, $p = 0.001$], whereas there were no differences between grammatical number (89.72%) and ungrammatical number (91.08%) [$F(1, 46) = 1.18$, $p = 0.283$]. The analyses by feature factor showed that participants were more accurate with grammatical person (92.61%) than with grammatical number (89.72%) [$F(1, 46) = 16.23$, $p < 0.001$], and they were significantly more accurate with ungrammatical person (95.73%) than with ungrammatical number (91.08%) [$F(1, 46) = 37.67$, $p < 0.001$].

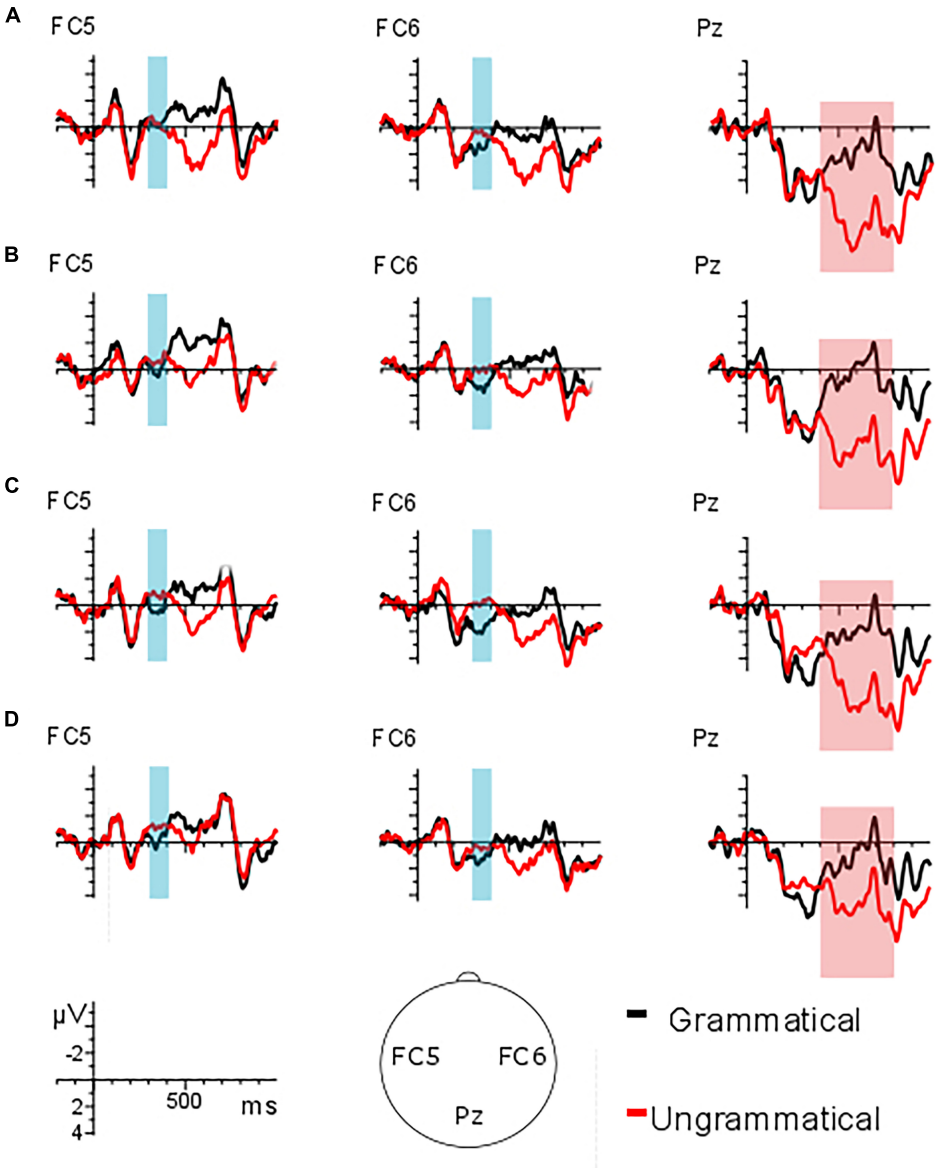

FIGURE 3 | (A) Person feature unaccusative predicate condition; **(B)** number feature unaccusative predicate condition; **(C)** person feature unergative predicate condition; **(D)** number feature unergative predicate condition.

Finally, a triple TYPE*GRAMMATICALITY*FEATURE interaction turned out to be significant [$F1(1, 46) = 9.28$, $p = 0.004$; $F2(1, 252) = 5.72$, $p = 0.017$]. The analyses by grammaticality factor showed that participants were accurate when performing the task with grammatical and ungrammatical unaccusatives containing person feature (94.01% vs. 95.36%) [$F(1, 46) = 2.32$, $p = 0.134$], and similarly with unaccusatives containing number feature (89.72% vs. 91.93%) [$F(1, 46) = 2.14$, $p = 0.15$]. In the unergative conditions participants were significantly more accurate with sentences containing ungrammatical person (96.1%) than with grammatical person (91.21%) [$F(1, 46) = 15.93$, $p < 0.001$], but no differences were found between grammatical (89.72%) and ungrammatical number (90.24%) [$F(1, 46) = 0.13$, $p = 0.716$]. The analyses by type factor revealed that participants were more accurate with sentences containing grammatical person feature in unaccusatives (94.01%) than in unergatives (91.21%) [$F(1, 46) = 9.41$, $p = 0.004$], whereas no differences were found between sentences containing ungrammatical person feature in unaccusatives (95.36%) and in unergatives (96.1%) [$F(1, 46) = 1.11$, $p = 0.296$]. With regard to number feature, no differences were found between grammatical unaccusative (89.72%) and (89.72%) unergative predicates [$F(1, 46) < 0.01$, $p = 1$], and neither between ungrammatical unaccusatives (91.93%) in contrast to ungrammatical unergative (90.24%) predicates [$F(1, 46) = 2.16$, $p = 0.148$]. Finally, the analyses by

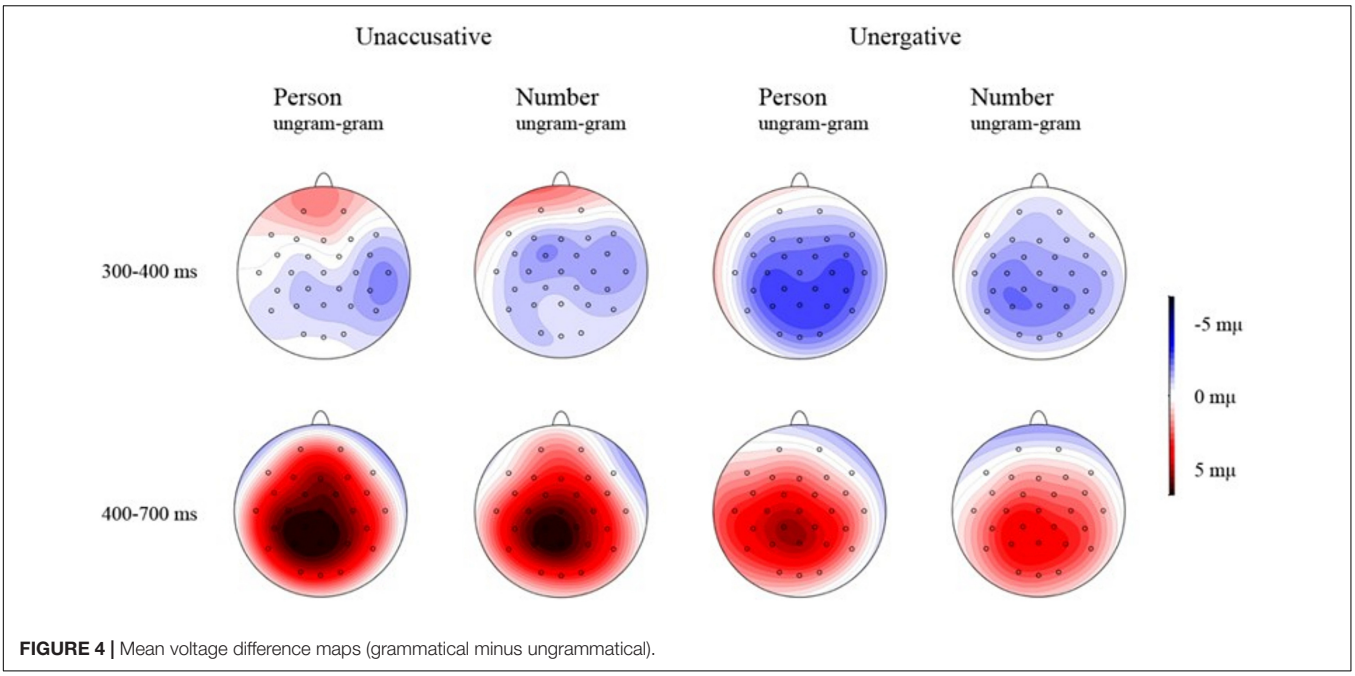

FIGURE 4 | Mean voltage difference maps (grammatical minus ungrammatical).

feature factor showed that participants were significantly more accurate with grammatical unaccusative sentences containing person feature (94.01%) than with number feature (89.72%) [$F(1, 46) = 22.79$, $p < 0.001$], and similarly ungrammatical unaccusative sentences containing person feature (95.36%) were judged more accurately than number feature (91.93%) [$F(1, 46) = 12.01$, $p = 0.001$]. Regarding unergative predicates, no differences were found between grammatical sentences containing person (91.21%) and number feature (89.72%) [$F(1, 46) = 2.18$, $p = 0.147$], but ungrammatical sentences containing person feature (96.1%) were judged significantly more accurately than sentences containing number feature (90.24%) [$F(1, 46) = 37.89$, $p < 0.001$].

The analysis of response times revealed a main TYPE effect [$F1(1, 46) = 18.21$, $p < 0.001$; $F2(1, 254) = 7.68$, $p < 0.006$] indicating that participants reacted faster to unaccusative predicates (637.29 ms) than to unergative predicates (673.08 ms). A main GRAMMATICALITY effect [$F1(1, 46) = 122.46$, $p < 0.001$; $F2(1, 252) = 453.86$, $p < 0.001$] revealed that participants were significantly faster reading ungrammatical sentences (565.16 ms) compared to their grammatical counterparts (745.22 ms). A FEATURE effect [$F1(1, 46) = 15.48$, $p < 0.001$; $F2(1, 254) = 11.22$, $p = 0.001$] revealed that participants were significantly faster reading sentences containing person feature (636.77 ms) than number feature (673.61 ms). A significant TYPE*GRAMMATICALITY interaction emerged [$F1(1, 46) = 5.04$, $p = 0.03$; $F2(1, 252) = 4.3$, $p = 0.039$]. The analyses by grammaticality factor showed that participants reacted faster to ungrammatical unaccusatives (557.11 ms) than to grammatical unaccusative (717.47 ms) predicates [$F1(1, 46) = 88.49$, $p < 0.001$; $F2(1, 252) = 190.11$, $p < 0.001$], and similarly participants responded faster to ungrammatical unergatives (573.21 ms) compared to their grammatical counterparts (772.96 ms) [$F1(1, 46) = 105.37$, $p < 0.001$; $F2(1, 252) = 274.53$, $p < 0.001$]. The analyses by type factor revealed significant differences between grammatical unaccusative and unergative predicates [$F1(1, 46) = 16.78$, $p < 0.001$; $F2(1, 252) = 0.02$, $p = 0.004$], indicating that participants reacted faster to grammatical unaccusatives (717.47 ms) than to grammatical unergatives (772.96 ms), but no differences were found between ungrammatical unaccusatives (557.11 ms) and ungrammatical unergative predicates (573.21 ms) [$F1(1, 46) = 2.45$, $p = 0.124$; $F2(1, 252) = 0.77$, $p = 0.3.25$].

Finally, a triple TYPE*GRAMMATICALITY*FEATURE interaction turned out to be marginally significant in the by subject analysis [$F(1, 46) = 3.25$, $p = 0.078$; $F2(1, 252) = 3.62$, $p = 0.058$]. The analyses by grammaticality factor showed that the unaccusative ungrammatical person condition (550.82 ms) was read faster than the grammatical person condition (705.5 ms) [$F(1, 46) = 60.54$, $p < 0.001$], and similarly for number (563.4 ms vs. 729.44 ms) [$F(1, 46) = 54.99$, $p < 0.001$]. In the unergative conditions participants were significantly faster with sentences containing ungrammatical person (532.75) than with grammatical person (757.99 ms) [$F(1, 46) = 109.43$, $p < 0.001$], and similarly they were faster with ungrammatical number (613.67 ms) than with grammatical number (787.93 ms) [$F(1, 46) = 59.2$, $p < 0.001$]. The analyses by type factor revealed participants were faster with sentences containing grammatical person feature in unaccusatives (705.5 ms) than in unergatives (757.99 ms) [$F(1, 46) = 8.4$, $p = 0.006$], whereas no differences emerged between ungrammatical person feature in unaccusatives (550.82 ms) and unergatives (532.75 ms) [$F(1, 46) = 1.23$, $p = 0.273$]. With regard to number feature, participants responded faster to grammatical sentences in unaccusatives (729.44 ms) than in unergatives (787.93 ms) [$F(1,$

TABLE 3 | Summary of the ERP results.

		300–400 ms		400–700 ms	
		Lateral	Midline	Lateral	Midline
	df	*F*	*F*	*F*	F
GRAM	1.23	***18.92	**11.13	***60.25	***59.63
TYPE	1.23	0.19	0.25	0.11	0.27
FEAT	1.23	0.01	1.26	**13.47	***20.94
TYPE*GRAM	1.23	2.67	4.9*	**9.34	**13.49
FEAT*GRAM	1.23	0.2	0.09	0.88	0.48
TYPE*FEAT*GRAM	1.23	0.83	0.36	0.21	2.04
GRAM*HEM	1.23	2.26	–	3.7	–
TYPE*GRAM*HEM	1.23	0.56	–	0.01	–
FEAT*GRAM*HEM	1.23	ˆ3.2	–	2.97	–
TYPE*FEAT*GRAM*HEM	1.23	0.14	–	1.35	–
GRAM*REGION	2.46	3.29	***14.42	***24.98	***56.36
TYPE*GRAM*REG	2.46	0.49	1.31	0.49	0.66
FEAT*GRAM*REG	2.46	0.3	0.99	0.69	0.29
TYPE*FEAT*GRAM*REG	2.46	0.43	1.2	0.71	0.21
GRAM*HEM*REG	2.46	0.7	–	1.45	–
TYPE*GRAM*HEM*REG	2.46	0.38	–	0.02	–
FEAT*GRAM*HEM*REG	2.46	0.33	–	0.05	–
TYPE*FEAT*GRAM*HEM*REG	2.46	0.14	–	1.13	–

*Main effects and interactions with grammaticality are shown. GRAM (grammaticality), TYPE (type), FEAT (feature), HEM (hemisphere), and REG (region). ˆp = < 0.1, *p = < 0.05, **p = < 0.01, ***p = < 0.001.*

46) = 10.73, $p = 0.002$], and similarly participants reacted faster to ungrammatical unaccusative (563.4 ms) than to ungrammatical unergatives (613.67 ms) [$F(1, 46) = 10.45$, $p = 0.002$]. Finally, the analyses by feature factor showed that participants reacted similarly to grammatical unaccusative sentences containing person feature (705.76 ms) and number feature (729.35) [$F(1, 46) = 1.62$, $p = 0.21$], and similarly there were no differences between ungrammatical unaccusative sentences containing person feature (550.81 ms) and number feature (561.63 ms) [$F(1, 46) = 0.75$, $p = 0.391$]. Regarding unergative predicates, no differences were found between grammatical sentences containing person (758.07 ms) and number feature (787.54 ms) [$F(1, 46) = 2.62$, $p = 0.113$], but ungrammatical sentences containing person feature (533.36 ms) were judged significantly faster than sentences containing number feature (612.96 ms) [$F(1, 46) = 25.48$, $p < 0.001$]. Overall, no differences between groups were observed in the behavioral measures.

ERP Results

Regarding the early time window (300–400 ms), the analysis of the lateral electrodes revealed a main GRAMMATICALITY effect [$F(1, 46) = 51.43$, $p < 0.001$] indicating a larger negativity for the ungrammatical conditions as compared to the grammatical ones (1 μV vs. 2.13 μV).

A significant FEATURE*GRAMMATICALITY interaction was found as well [$F(1, 46) = 4.5$, $p = 0.039$]. Further analysis (by grammaticality) showed a significantly larger negativity for the ungrammatical person condition (0.75 μV) in comparison to the grammatical one (2.15 μV) [$F(1, 46) = 42.5$, $p < 0.001$] and a larger negativity for the ungrammatical number condition (1.24 μV) in comparison to the grammatical number condition (2.11 μV) [$F(1, 46) = 20.38$, $p < 0.001$]. The comparison by feature revealed no differences between the grammatical person (2.15 μV) and number feature (2.11 μV) conditions [$F(1, 47) = 0.04$, $p = 0.841$], but it revealed a larger negativity for the ungrammatical person manipulations (0.75 μV) in comparison to the number manipulations (1.24 μV) [$F(1, 46) = 6.51$, $p = 0.014$].

Regarding the midline electrodes, a main effect of GRAMMATICALITY showed that overall ungrammatical conditions (2.35 μV) displayed a larger negativity than grammatical conditions (3.29 μV) [$F(1, 46) = 20.43$, $p > 0.001$].

A significant TYPE*GRAM*GROUP interaction [$F(1, 46) = 6.4$, $p = 0.015$] showed (by grammaticality factor) that natives revealed a larger negativity for the ungrammatical unaccusative condition (2.46 μV) than for the grammatical unaccusative condition (3.66 μV) [$F(1, 46) = 13.5$, $p = 0.001$] and also a larger negativity for the ungrammatical unergative condition (2.84 μV) than for the grammatical unergative condition (3.62 μV) [$F(1, 46) = 4.48$, $p = 0.040$], whereas non-natives only elicited a larger negativity for the ungrammatical unergative condition (1.7 μV) compared to the grammatical unergative condition (3.07 μV) [unaccusative: $F(1, 46) = 1.76$, $p = 0.191$; unergative: $F(1, 46) = 13.36$, $p = 0.001$]. In the analysis by type factor, no differences were found between grammatical unaccusative and unergative conditions neither in natives [$F(1, 46) = 0.2$, $p = 0.889$] nor in non-natives [$F(1, 46) = 0.71$, $p = 0.405$], and similarly no differences were found between ungrammatical unaccusative and unergative conditions neither in natives [$F(1, 46) = 0.95$, $p = 0.334$] nor in non-natives [$F(1, 46) = 2.67$, $p = 0.109$]. The *T*-test showed that the negativity elicited in non-natives was marginally larger than in natives for the unergative ungrammatical condition (1.7 μV vs. 2.84 μV) [$F(1, 46) = 0.45$, $p = 0.09$].

Regarding the 400–700 ms time window, the analysis of the lateral electrodes revealed a main GRAMMATICALITY effect [$F(1, 46) = 103.8$, $p < 0.001$] indicating a larger positivity for the ungrammatical conditions as compared to the grammatical ones (2.33 μV vs. 0.08 μV). A significant main effect of FEATURE also emerged [$F(1, 46) = 13.14$, $p = 0.001$], indicating that overall person feature generated a larger positivity as compared to number feature (1.48 μV vs. 0.93 μV).

A significant TYPE*GRAMMATICALITY interaction was found [$F(1, 46) = 6.1$, $p = 0.017$]. Further analysis (by grammaticality) showed a significantly larger positivity for the ungrammatical unaccusative condition (2.54 μV) in comparison to the grammatical one (0.02 μV) [$F(1, 46) = 108.41$, $p < 0.001$] and a larger positivity for the ungrammatical unergative condition (2.13 μV) in comparison to the grammatical number condition (0.13 μV) [$F(1, 46) = 65.39$, $p < 0.001$]. The comparison by type revealed no differences between the grammatical unaccusative (0.02 μV) and unergative (0.13 μV) conditions [$F(1, 46) = 0.4$, $p = 0.529$], but a larger positivity emerged for ungrammatical unaccusative manipulations (2.54 μV) in comparison to the ungrammatical unergative

manipulations (2.13 μV) [$F(1, 46) = 4.52$, $p = 0.039$]. A significant TYPE*GRAM*REGION [$F(2, 46) = 5.82$, $p = 0.012$] interaction also showed a larger positivity for ungrammatical conditions in comparison to grammatical conditions in unaccusatives [frontal: $F(1, 46) = 31.78$, $p < 0.001$; central: $F(1, 46) = 127.11$, $p < 0.001$; posterior: $F(1, 46) = 123.52$, $p < 0.001$] and in unergatives [frontal: $F(1, 46) = 27.37$, $p < 0.001$; central: $F(1, 46) = 76.9$, $p < 0.001$; posterior: $F(1, 46) = 54.44$, $p < 0.001$] in all three regions. The comparison by type revealed no differences between grammatical conditions, but it revealed significant differences between ungrammatical unaccusative and ungrammatical unergative conditions in central [$F(1, 46) = 5.31$, $p = 0.026$] and posterior [$F(1, 46) = 5.2$, $p = 0.027$] electrodes, thus indicating that ungrammatical unaccusatives elicit a larger positivity than ungrammatical unergatives.

Regarding the midline electrodes, a main effect of FEATURE [$F(1, 46) = 21.85$, $p < 0.001$] showed that overall person (3.01 μV) displayed a larger positivity than number (2.15 μV). A further main effect of GRAMMATICALITY [$F(1, 46) = 120.16$, $p > 0.001$] revealed that overall ungrammatical conditions (4.28 μV) displayed a larger positivity than grammatical conditions (0.87 μV).

A FEATURE*GROUP interaction [$F(1, 46) = 6.88$, $p = 0.012$] emerged. The analysis by feature factor revealed no differences between person and number regarding natives [$F(1, 46) = 2.1$, $p = 0.154$], but it showed that person elicited a larger positivity than number in non-natives (2.78 μV vs. 1.44 μV) [$F(1, 46) = 26.63$, $p < 0.001$]. The *t*-test showed that there were no group differences regarding person feature (3.23 μV vs. 2.78 μV) [$F(1, 46) = 0.46$, $p = 0.505$], but natives elicited a larger positivity than non-natives with regard to number feature (2.86 μV vs. 1.44 μV) [$F(1, 46) = 0.02$, $p = 0.022$].

A TYPE*GRAM interaction was also found [$F(1, 46) = 9.19$, $p = 0.004$]. Further analysis by grammaticality showed that ungrammatical unaccusatives (4.64 μV) elicited a larger positivity than grammatical unaccusatives (0.8 μV) [$F(1, 46) = 112.21$, $p > 0.001$], and similarly ungrammatical unergatives elicited a larger positivity (3.92 μV) than grammatical unergatives (0.94 μV) [$F(1, 46) = 81.51$, $p > 0.001$]. Analysis by type showed that there are no differences between grammatical unaccusatives (0.8 μV) and grammatical unergatives (0.94 μV) [$F(1, 46) = 0.36$, $p = 0.55$], whereas ungrammatical unaccusatives (4.64 μV) elicited a larger positivity than ungrammatical unergatives (3.92 μV) [$F(1, 46) = 7.5$, $p = 0.009$].

A marginally significant TYPE*GRAM*GROUP interaction [$F(1, 46) = 4.02$, $p = 0.051$] showed (by grammaticality factor) that both natives and non-natives revealed a larger positivity for the ungrammatical unaccusative condition than for the grammatical unaccusative condition [natives: $F(1, 46) = 60.29$, $p < 0.001$; non-natives: $F(1, 46) = 50.3$, $p < 0.001$] and also a larger positivity for the ungrammatical unergative condition than for the grammatical unergative condition [natives: $F(1, 46) = 69.71$, $p < 0.001$; non-natives: $F(1, 46) = 25.21$, $p < 0.001$]. In the analysis by type factor, no differences were found between the grammatical unaccusative and unergative conditions neither in natives [$F(1, 46) = 0.62$, $p = 0.435$] nor in non-natives [$F(1, 46) = 2.75$, $p = 0.104$], and no differences were found between ungrammatical unaccusative and unergative conditions in natives either [$F(1, 46) = 2.16$, $p = 0.149$]. However, in non-natives ungrammatical unaccusatives elicited a larger positivity than ungrammatical unergatives (4.02 μV vs. 3.14 μV) [$F(1, 46) = 5.67$, $p = 0.021$]. The *t*-test revealed that natives elicited a marginally larger positivity than non-natives in the unergative ungrammatical condition (4.71 μV vs. 3.14 μV) [$F(1, 46) = 1.98$, $p = 0.054$].

There was also a significant TYPE*FEATURE*GRAM interaction [$F(2, 46) = 7.22$, $p = 0.01$]. The analysis by grammaticality factor showed that the positivity elicited by the ungrammatical sentences was significantly larger than that yielded by the grammatical sentences in all the conditions [unaccusative person: $F(1, 23) = 75.1$, $p < 0.001$; unaccusative number: $F(1, 23) = 100.22$, $p < 0.001$; unergative person: $F(1, 23) = 93.01$, $p < 0.001$; unergative number: $F(1, 23) = 36.42$, $p < 0.001$]. The analysis by type factor revealed no differences across predicate type regarding person feature [grammatical person: $F(1, 23) = 1.11$, $p < 0.298$; ungrammatical person: $F(1, 23) = 0.05$, $p = 0.828$]. Nevertheless, regarding number feature grammatical unergatives elicited a larger positivity than grammatical unaccusatives (0.83 μV vs. 0.22 μV) [grammatical number: $F(1, 23) = 5.81$, $p = 0.02$], and ungrammatical unaccusatives elicited a larger positivity than ungrammatical unergatives (4.24 μV vs. 3.29 μV) [$F(1, 23) = 8.36$, $p = 0.006$]. Concurrently, the analysis by feature factor showed that person elicited a larger positivity than number feature in the grammatical (1.38 μV vs. 0.22 μV) and ungrammatical (5.05 μV vs. 4.24 μV) unaccusative condition [grammatical: $F(1, 23) = 15.07$, $p < 0.001$; ungrammatical: $F(1, 23) = 4.92$, $p = 0.031$]. Regarding unergatives, no differences were found between person and number in the grammatical condition [$F(1, 23) = 0.34$, $p = 0.563$], but a significant difference was found between person and number in the ungrammatical condition, showing that both natives and non-natives generate a larger positivity when processing ungrammatical person unergatives (4.55μV) than ungrammatical number unergatives (3.3μV) [$F(1, 23) = 31.32$, $p < 0.001$].

Finally, a marginally significant GRAM*REGION*GROUP interaction also emerged [$F(2, 46) = 3.19$, $p = 0.056$]. The analysis by grammaticality showed that both natives and non-natives elicited a larger positivity in ungrammatical sentences than in their grammatical counterparts in frontal electrodes [natives: $F(1, 46) = 34.78$, $p > 0.001$; non-natives: $F(1, 46) = 25.49$, $p > 0.001$], central electrodes [natives: $F(1, 46) = 70.04$, $p > 0.001$; non-natives: $F(1, 46) = 39.1$, $p > 0.001$], and posterior electrodes [natives: $F(1, 47) = 94.26$, $p > 0.001$; non-natives: $F(1, 46) = 51.57$, $p > 0.001$]. The analysis by group showed that no differences obtained between native and non-natives' grammatical sentences in frontal electrodes [$F(1, 46) = 0.55$, $p = 0.284$], nor central electrodes [$F(1, 46) = 0.35$, $p = 0.523$], nor posterior electrodes [$F(1, 46) = 0.14$, $p = 0.484$]. Similarly, no differences between native and non-natives obtained in ungrammatical sentences in frontal electrodes [$F(1, 46) = 0.44$, $p = 0.177$] and in central electrodes [$F(1, 46) = 0.72$, $p = 0.114$], but natives elicited a marginally larger positivity in posterior electrodes compared to

TABLE 4 | Summary of the ERP results.

		300–400 ms		400–700 ms	
		Lateral	Midline	Lateral	Midline
	df	*F*	*F*	*F*	*F*
GROUP	1.23	0.01	1.42	0.81	2.36
GRAM	1.23	***51.43	***20.43	***103.8	***120.61
GRAM*GROUP	1.23	1.79	0.06	0.47	2.18
TYPE	1.23	0.0	0.0	0.86	2.02
TYPE*GROUP	1.23	0.62	0.64	1.3	0.32
FEAT	1.23	2.65	0.25	**13.14	***21.85
FEAT*GROUP	1.23	^3.07	^3.96	^3.55	*6.88
TYPE*GRAM	1.23	0.69	0.84	**6.1	**9.19
TYPE*GRAM*GROUP	1.23	2.26	*6.4	1.62	^4.02
FEAT*GRAM	1.23	*4.5	2.21	0.06	1.17
FEAT*GRAM*GROUP	1.23	1.87	0.97	0.96	0.01
TYPE*FEAT*GRAM	1.23	2.53	1.84	2.47	*7.22
TYPE*FEAT*GRAM*GROUP	1.23	0.11	0.3	1.07	0.79
GRAM*HEM	1.23	0.66	–	1.58	–
GRAM*HEM*GROUP	1.23	1.11	–	1.35	–
TYPE*GRAM*HEM	1.23	0.83	–	0.74	–
TYPE*GRAM*HEM*GROUP	1.23	^3.41	–	0.96	–
FEAT*GRAM*HEM	1.23	0.03	–	0.11	–
FEAT*GRAM*HEM*GROUP	1.23	^2.94	–	^3.26	–
TYPE*FEAT*GRAM*HEM	1.23	0.32	–	^2.93	–
TYPE*FEAT*GRAM*HEM*GROUP	1.23	0.0	–	0.0	–
GRAM*REGION	2.46	**10.82	***13.98	***41.46	***77.1
GRAM*REGION*GROUP	2.46	3.13	0.48	0.34	^3.19
TYPE*GRAM*REG	2.46	3.03	1.28	*5.82	0.91
TYPE*GRAM*REG*GROUP	2.46	0.82	0.08	2.62	0.83
FEAT*GRAM*REG	2.46	1.1	^2.88	1.62	0.53
FEAT*GRAM*REG*GROUP	2.46	0.14	0.43	0.1	0.37
TYPE*FEAT*GRAM*REG	2.46	0.3	2.05	1.38	2.4
TYPE*FEAT*GRAM*REG*GROUP	2.46	0.22	0.02	0.26	1.02
GRAM*HEM*REG	2.46	*5.15	–	0.7	–
GRAM*HEM*REG*GROUP	2.46	1.84	–	0.53	–
TYPE*GRAM*HEM*REG	2.46	0.01	–	0.19	–
TYPE*GRAM*HEM*REG*GROUP	2.46	0.52	–	0.15	–
FEAT*GRAM*HEM*REG	2.46	0.27	–	0.3	–
FEAT*GRAM*HEM*REG*GROUP	2.46	0.23	–	0.62	–
TYPE*FEAT*GRAM*HEM*REG	2.46	0.03	–	0.73	–
TYPE*FEAT*GRAM*HEM*REG*G	2.43	0.28	–	0.31	–

Main effects and interactions with grammaticality are shown.
*GRAM (grammaticality), TYPE (type), FEAT (feature), HEM (hemisphere) and REG (region). ^p = < 0.1, *p = < 0.05, **p = < 0.01, ***p = < 0.001.*

non-natives (6.82 μV vs. 4.97 μV) [$F(1, 46) = 2.19$, $p = 0.065$] (See **Table 4** for the summary of the results).

DISCUSSION

The present study aimed to examine whether native-like processing can be achieved in a second language when the linguistic features tested are shared by L1 and L2. To this purpose we tested early and high proficient non-native speakers of Basque while processing intransitive predicates (unergatives/unaccusatives) and phi-features (person/number). These results were compared to those previously obtained from native speakers in the same tasks (Martinez de la Hidalga et al., 2019).

Overall, these early and proficient non-native speakers were indistinguishable from natives: (a) in the acceptability task, non-natives were faster and more accurate in the unaccusative condition and in the person violation condition; (b) they displayed a larger positivity for unaccusative violations than for unergative violations; (c) in the unergative condition, they displayed a larger positivity for person than for number feature violations; (d) number violations elicited larger positivity in the unaccusative condition than in the unergative condition.

In Martinez de la Hidalga et al. (2019), we found differences in the processing of unaccusative and unergative predicates, thus supporting the Unaccusative Hypothesis. Nevertheless, we showed that in Basque, contrary to the predictions made by the Unaccusative Hypothesis, unaccusative predicates are not costlier to process than unergative predicates. We thus provided new evidence in support of the view which advocates for no syntactic movement for subjects of unaccusatives in Basque (Laka, 2006a,b; Levin, 1983). Evidence for greater processing costs for unaccusative predicates compared to unergative predicates have been repeatedly found in nominative-accusative languages (Bastiaanse and van Zonneveld, 2005; Friedmann et al., 2008; Koring et al., 2012; Meltzer-Asscher et al., 2015; Dekydtspotter and Seo, 2017; *inter alia*). One possibility for Spanish bilinguals may have been to show traces of the processing of their native language (nominative-accusative) when processing intransitive predicates in Basque (ergative-absolutive), where no extra processing costs are found for unaccusative predicates. Instead, early and high proficient Spanish-Basque bilinguals processed intransitive predicates as do natives, and displayed measures of greater processing costs for unergatives than for unaccusatives.

Although by and large L2 speakers behaved native-like, their electrophysiological activity revealed a few minor differences: (a) non-natives did not generate a N400 in response to unaccusative violations; (b) the P600 generated by violations in the unergative condition was smaller than that generated by natives; (c) regarding phi-features, non-natives generated smaller negativity for number than natives.

Regarding (a) a lack of negativity in non-natives has been often reported in the literature (Hagoort et al., 1993; Osterhout et al., 1996; Münte et al., 1997; Hagoort and Brown, 2000; Alemán-Bañón and Rothman, 2019; *i.a.*). According to Hahne (2001) smaller or absent negativity in non-native speakers when detecting ungrammaticality may be due to a reduced degree of automaticity in the activation of processing resources.

Regarding (b) quantitative differences in the P600 have also been reported for non-native speakers and are usually attributed to differences in the frequency of use (Osterhout et al., 2006; Rossi et al., 2006). We cannot discard the possibility that differences in the frequency of use had an effect in the processing of intransitive predicate. This factor shall be taken into consideration in future studies in order to discard this possibility.

Besides, there is a morphological difference between the native and non-native language: the processing of unergative subject-verb agreement in Basque involves the processing of ergative case, a case marker not existing in Spanish. A speculative explanation for the slight enhancement of the N400 for unergatives and the decrease of the P600 for non-natives could be that non-natives displayed a smaller sensitivity for processing a case marking not present in their native language, compared to unaccusatives, where case is morphologically unmarked.

Finally, regarding (c) phi features, non-natives generated smaller negativity for number than natives. In any case, the effect and tendency to generate larger negativity for person than for number was the same for native and non-native speakers.

To conclude, in the present study we provided evidence that native-like processing is attainable for early and proficient bilinguals whenever the linguistic properties are shared by their two languages, as suggested by the LDH. We, therefore, argue that cross-linguistic similarity is an important factor that deserves further consideration to better understand what drives bilinguals to process a second language *native-like*.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee for Research involving human beings at the University of the Basque Country (UPV/EHU). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GM: data acquisition, analysis and writing. AZ: conceptualization, experimental design, assistance with data analysis, and writing. IL: conceptualization, experimental design and writing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the Spanish Ministerio de Ciencia e Innovación (FPI-2017-BES-2016-076456) (GM), Ministerio de Ciencia, Innovación y Universidades (PID2019-104016GB-I00), and the Basque Government (IT1169-19).

REFERENCES

Alemán-Bañón, J., and Rothman, J. (2019). Being a participant matters: event-related potentials show that markedness modulates person agreement in Spanish. *Front. Psychol.* 10:746. doi: 10.3389/fpsyg.2019.00746

Bastiaanse, R., and van Zonneveld, R. (2005). Sentence production with verbs of alternating transitivity in agrammatic Broca's aphasia. *J. Neurolinguist.* 18, 57–66. doi: 10.1016/j.jneuroling.2004.11.006

Brice, H., Mencl, W. E., Frost, S. J., Bick, A. S., Rueckl, J. G., Pugh, K. R., et al. (2019). Neurobiological signatures of L2 proficiency: evidence from a bi-directional cross-linguistic study. *J. Neurolinguist.* 50, 7–16.

Burzio, L. (1986). *Italian Syntax: A Government-Binding Approach*. Dordrecht: Springer.

Caffarra, S., Molinaro, N., Davidson, D., and Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: an empirical review. *Neurosci. Biobehav. Rev.* 51C, 31–47. doi: 10.1016/j.neubiorev.2015.01.010

Carminati, M. N. (2005). Processing reflexes of the feature hierarchy (Person > Number > Gender) and implications for linguistic theory. *Lingua* 115, 259–285. doi: 10.1016/j.lingua.2003.10.006

Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psychol.* 27, 3–42. doi: 10.1017/S0142716406060024

Dekydtspotter, L., and Seo, H.-K. (2017). Transitivity in the processing of intransitive clauses: a category-based prediction in low-intermediate learners of english. *Stud. Sec. Lang. Acquisit.* 39, 527–552. doi: 10.1017/s0272263116000188

Díaz, B., Erdocia, K., de Menezes, R. F., Mueller, J. L., Sebastián-Gallés, N., and Laka, I. (2016). Electrophysiological correlates of second-language syntactic processes are related to native and second language distance regardless of age of acquisition. *Front. Psychol.* 7:133. doi: 10.3389/fpsyg.2016.00133

Díaz, B., Sebastián-Gallés, N., Erdocia, K., Mueller, J. L., and Laka, I. (2011). On the cross-linguistic validity of electrophysiological correlates of morphosyntactic processing: a study of case and agreement violations in Basque. *J. Neurolinguist.* 24, 357–373. doi: 10.1016/j.jneuroling.2010.12.003

Erdocia, K., Zawiszewski, A., and Laka, I. (2014). Word order processing in a second language: from VO to OV. *J. Psycholinguist. Res.* 43, 815–837. doi: 10.1007/s10936-013-9280-4

Foucart, A., and Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *J. Mem. Lang.* 66, 226–248. doi: 10.1016/j.jml.2011.07.007

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 529–534. doi: 10.1073/pnas.012611199

Friedmann, N., Taranto, G., Shapiro, L. P., and Swinney, D. (2008). The leaf fell (the leaf): the online processing of unaccusatives. *Linguist. Inq.* 39, 355–377. doi: 10.1162/ling.2008.39.3.355

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9

Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24, 95–112. doi: 10.1007/BF02289823

Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (sps) as an ERP measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *J. Psycholinguist. Res.* 30, 251–266.

Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (2018). A critical period for second language acquisition: evidence from 2/3 million english speakers. *Cognition* 177, 263–277. doi: 10.1016/j.cognition.2018.04.007

Hartsuiker, R. J., Pickering, M. J., and Veltkamp, E. (2004). Is syntax separate or shared between languages? *Psychol. Sci.* 15, 409–414. doi: 10.1111/j.0956-7976.2004.00693.x

Hernandez, A. E., Hofmann, J., and Kotz, S. A. (2007). Age of acquisition modulates neural activity for both regular and irregular syntactic functions. *NeuroImage* 36, 912–923.

Hernandez, A., Li, P., and MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends Cogn. Sci.* 9, 220–225. doi: 10.1016/j.tics.2005.03.003

Koring, L., Mak, P., and Reuland, E. (2012). The time course of argument reactivation revealed: using the visual world paradigm. *Cognition* 123, 361–379. doi: 10.1016/j.cognition.2012.02.011

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain Lang.* 109, 68–74.

Kotz, S. A., Holcomb, P. J., and Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in early bilinguals and monolinguals. *Acta Psychol.* 128, 514–527. doi: 10.1016/j.actpsy.2007.10.003

Laka, I. (2006a). Deriving split ergativity in the progressive. *Ergativity* 65, 173–195.

Laka, I. (2006b). "On the nature of case in Basque: structural or inherent?," in *Organizing Grammar*, eds H. Broekhuis, N. Corver, R. Huijbregts, U. Kleinhenz, and J. Koster (Berlin: Walter de Gruyter).

Lenneberg, E. H. (1967). The biological foundations of language. *Hosp. Pract.* 2, 59–67.

Levin, B. (1983). *On the Nature of Ergativity*. [Doctoral dissertation]. Cambridge, MA: Massachussets Institute of Technology.

MacWhinney, B. (2005). Extending the competition model. *Int. J. Bilingualism* 9, 69–84. doi: 10.1177/13670069050090010501

Mancini, S. (2018). *Features and Processing in Agreement*. Cambridge, MA: Cambridge Scholars Publishing.

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject–verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055

Martinez de la Hidalga, G., Zawiszewski, A., and Laka, I. (2019). Eppur non si muove: experimental evidence for the Unaccusative Hypothesis and distinct Φ-feature processing in Basque. *Glossa* 4:120. doi: 10.5334/gjgl.829

Matzke, M., Mai, H., Nager, W., Rüsseler, J., and Münte, T. (2002). The costs of freedom: an ERP – study of non-canonical sentences. *Clin. Neurophysiol.* 113, 844–852. doi: 10.1016/s1388-2457(02)00059-7

Meltzer-Asscher, A., Mack, J. E., Barbieri, E., and Thompson, C. K. (2015). How the brain processes different dimensions of argument structure complexity: evidence from fMRI. *Brain Lang.* 142, 65–75. doi: 10.1016/j.bandl.2014.12.005

Montrul, S. (2005). On knowledge and development of unaccusativity in Spanish L2 acquisition. *Linguistics* 43, 1153–1190.

Münte, T. F., Matzke, M., and Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *J. Cogn. Neurosci.* 9, 318–329. doi: 10.1162/jocn.1997.9.3.318

Ojima, S., Nakata, H., and Kakigi, R. (2005). An ERP study of second language learning after childhood: effects of proficiency. *J. Cogn. Neurosci.* 17, 1212–1228.

Ortiz de Urbina, J. (1989). *Parameters in the Grammar of Basque: A GB Approach to Basque Syntax*. Dordrecht: Foris.

Oshita, H. (2001). The unaccusative trap in second language acquisition. *Stud. Sec. Lang. Acquisit.* 23, 279–304. doi: 10.1017/s0272263101002078

Osterhout, L., McKinnon, R., Bersick, M., and Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: is the syntactic positive shift a member of the p300 family? *J. Cogn. Neurosci.* 8, 507–526. doi: 10.1162/jocn.1996.8.6.507

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., and Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: a means for exploring the neurocognition of second language processing. *Lang. Learn.* 56, 199–230.

Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. *Annu. Meet. Berkeley Linguist. Soc.* 4, 157–190. doi: 10.3765/bls.v4i0.2198

Rossi, S., Gugler, M. F., Friederici, A. D., and Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *J. Cogn. Neurosci.* 18, 2030–2048. doi: 10.1162/jocn.2006.18.12.2030

Schwartz, B., and Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Sec. Lang. Res.* 12, 40–72.

Steinhauer, K., White, E. J., and Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Lang. Res.* 25, 13–41. doi: 10.1177/0267658308098995

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

Vandenberghe, B., Perez, M. M., Reynvoet, B., and Desmet, P. (2019). The role of Event-Related Potentials (ERPs) as sensitive measures in L2 vocabulary acquisition research. *J. Eur. Sec. Lang. Assoc.* 3, 35–45. doi: 10.22599/jesla.60

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., and Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159–170.

Weber-Fox, C. M., and Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *J. Cogn. Neurosci.* 8, 231–256. doi: 10.1162/jocn.1996.8.3.231

Yuan, B. (1999). Acquiring the unaccusative unergative distinction in a second language: evidence from English-speaking learners of L2 Chinese. *Linguistics* 37, 275–296.

Zawiszewski, A., Gutiérrez, E., Fernández, B., and Laka, I. (2011). Language distance and non-native syntactic processing: evidence from event-related potentials. *Bilingualism Lang. Cogn.* 14, 400–411. doi: 10.1017/s1366728910000350

Zawiszewski, A., and Laka, I. (2020). Bilinguals processing noun morphology: evidence for the language distance hypothesis from event-related potentials. *J. Neurolinguist.* 55:100908. doi: 10.1016/j.jneuroling.2020.100908

Zawiszewski, A., Santesteban, M., and Laka, I. (2016). Phi-features reloaded: an event-related potential study on person and number agreement processing. *Appl. Psycholinguist.* 37, 601–626. doi: 10.1017/s014271641500017x

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fpsyg.2021.716485

A Challenge for Contrastive L1/L2 Corpus Studies: Large Inter- and Intra-Individual Variation Across Morphological, but Not Global Syntactic Categories in Task-Based Corpus Data of a Homogeneous L1 German Group

Anna Shadrova, Pia Linscheid, Julia Lukassek, Anke Lüdeling and Sarah Schneider*

Department of German Studies and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany

OPEN ACCESS

Edited by:
Pedro Guijarro-Fuentes, University of the Balearic Islands, Spain

Reviewed by:
Erin Conwell, North Dakota State University, United States
Amanda Edmonds, Université Côte d'Azur, France

***Correspondence:**
Anna Shadrova
anna.shadrova@hu-berlin.de

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

Received: *28 May 2021*
Accepted: *21 October 2021*
Published: *25 November 2021*

Citation:
Shadrova A, Linscheid P, Lukassek J, Lüdeling A and Schneider S (2021) A Challenge for Contrastive L1/L2 Corpus Studies: Large Inter- and Intra-Individual Variation Across Morphological, but Not Global Syntactic Categories in Task-Based Corpus Data of a Homogeneous L1 German Group. Front. Psychol. 12:716485. doi: 10.3389/fpsyg.2021.716485

In this paper, we present corpus data that questions the concept of native speaker homogeneity as it is presumed in many studies using native speakers (L1) as a control group for learner data (L2), especially in corpus contexts. Usage-based research on second and foreign language acquisition often investigates quantitative differences between learners, and usually a group of native speakers serves as a control group, but often without elaborating on differences within this group to the same extent. We examine inter-personal differences using data from two well-controlled German native speaker corpora collected as control groups in the context of second and foreign language research. Our results suggest that certain linguistic aspects vary to an extent in the native speaker data that undermines general statements about quantitative expectations in L1. However, we also find differences between phenomena: while morphological and syntactic sub-classes of verbs and nouns show great variability in their distribution in native speaker writing, other, coarser categories, like parts of speech, or types of syntactic dependencies, behave more predictably and homogeneously. Our results highlight the necessity of accounting for inter-individual variance in native speakers where L1 is used as a target ideal for L2. They also raise theoretical questions concerning a) explanations for the divergence between phenomena, b) the role of frequency distributions of morphosyntactic phenomena in usage-based linguistic frameworks, and c) the notion of the individual adult native speaker as a general representative of the target language in language acquisition studies or language in general.

Keywords: corpus linguistic analysis, quantitative linguistics, morphology, usage-based linguistics, verb morphology, noun morphology, language variation and corpus

1. INTRODUCTION

The frequency of occurrence of linguistic elements and categories, such as words, clause types, morphological, syntactic, or lexical features, has played a central role in usage-based linguistics (Ellis, 2002; Granger, 2005, 2015; Goldberg, 2006, 2013; Biber and Jones, 2009; Paquot and Granger, 2012; Zeldes, 2012; Bybee, 2013; Gries, 2013, 2014; Hirschmann et al., 2013; Bestgen and Granger, 2014; Gries and Ellis, 2015; Hirschmann, 2015; Diessel and Hilpert, 2016, among many others). In connectivist models of learning and acquisition, linguistic ability is modeled as the result of entrenchment of neuronal pathways through repeated exposure. Frequency is a crucial factor in deciding which combinations or connections emerge and persist (Croft, 2000; Tomasello, 2000, 2009; Bybee and Hopper, 2001; Goldberg et al., 2004; Schmitt, 2004; Gries and Wulff, 2005; Hoey, 2005; Ellis, 2006, 2012; Divjak and Caldwell-Harris, 2015; Ellis and Wulff, 2015, and many others). Since language learners are overall less exposed to target language input compared to native speakers, frequency has also served as an explanation for divergent degrees of language attainment in second language acquisition (SLA), for instance in studies that work with the concepts of over- and underuse (Paquot and Granger, 2012; Bestgen and Granger, 2014, and others), especially in connection with Contrastive Interlanguage Analysis, an influential method in learner corpus research (Granger, 2015). Deviations from native speaker frequencies found in corpora are frequently interpreted as evidence for true differences between native speakers and learners rather than random fluctuation. Typically, cumulative corpus counts or relative frequencies normalized to corpus size or a fixed number of tokens, such as one million words, are used for this.

This can be problematic, since native speakers are not monolithic in their use of language, as has been studied explicitly in variationist and socio-linguistic approaches (Eckert, 2016; Szmrecsanyi, 2017; Bayley, 2019), including SLA, e.g., Linford et al. (2016) and Gurzynski-Weiss et al. (2018). Careful analysis of inter-individual variability in L1 in quantitative learner corpus studies remains rare, as has been pointed out by Gries and Deshors (2014)[1]. This lack of attention can in part be attributed to limitations of the data as it tends to occur in corpora. Individual texts frequently contain only few instances of the categories of interest, especially where lexical or phraseological material is concerned (Shadrova, 2020, chap. 4), and thus often do not allow for a meaningful analysis of inter-individual variance. To a degree, this is unavoidable, since corpus data is not as neatly controllable with respect to the elicitation of linguistic features as some experimental data, and some features do not occur frequently unless prompted directly. Limited ability to consider inter- or intra-individual differences can also be due to corpus design, especially where data is collected without attribution to individuals (like web corpora) or texts in the data differ too much in text length, type, or genre to be easily comparable (like homework corpora collected over years). However, importantly, this common practice in quantitative corpus linguistics is also an extension of the underlying philosophy held by most models of language acquisition in usage-based linguistics—if frequency is modeled as somewhat stable in the in- and output, *there should be no problem* with cumulative data.

The data we present in this study suggests that this may not be true on all levels of granularity. Our two corpora of essays written by German native speakers, Falko (Reznicek et al., 2012) and Kobalt (Zinsmeister et al., 2012)[2], were collected with the aim of maximally homogenizing the data regarding age, environment, and conditions of elicitation and prompt (intended to elicit homogeneous topic, register, and genre), as they were originally compiled as control group data in a contrastive L1/L2 paradigm[3].

In spite of this maximally homogeneous composition, we find surprisingly high levels of inter-individual variation in the distribution of morphological categories of verbs and nouns and syntactic subclasses of verbs. At the same time, we find high convergence between participants regarding the distributions of global syntactic categories (parts of speech and syntactic dependencies). The purpose of this paper is to present and discuss these differences and similarities, and to highlight some of the repercussions of these findings on usage-based theory and corpus methodology. It is, to our best knowledge, the first corpus-based and quantitative account of both morphological and syntactic categories in homogeneous corpora of German. While we enter the discussion from a learner corpus perspective, we will not discuss learner data in this paper in order to give space for a discussion of what is designed as control group data. We argue that converging frequency distributions cannot be expected across levels of granularity even in socially and functionally highly homogeneous data. Rather, it appears that distributions converge on *some*, but not *all* linguistic levels. It follows that cumulative corpus accounts can be grossly misleading depending on the phenomenon they wish to investigate. They do not account for the full complexity of native speaker writing and may lead to over- or underestimations or incomplete models of true differences between L1 and L2.

The two stances—that cumulative data can be of sufficiently fine resolution and that native speakers can vary in their linguistic expression—are in principle not contradictory. Linguistic variationism, as we understand it, focuses on social, situational, or linguistic, but always *functional*, i.e., *stratified*, variability cf. Eckert (2016), Bayley (2019), and Szmrecsanyi (2019). This is expressed in a matrix of variants of a variable by factor, such as group membership (e.g., an individual's belonging to a certain age bracket, geographic area, cultural background, etc., cf. Dubois and Sankoff, 2001; Lüdeling, 2017; Szmrecsanyi, 2017); groups formed from less transparently available traits, such as aptitude, motivation; or functional variance, such as situational

[1]One notable exception is Mulder and Hulstijn (2011).

[2]The data is described in detail in section 3. Annotation details are provided in the **Appendix** in **Supplementary Materials**.

[3]One could argue that it is next to impossible to keep the genre, register, or broader function of text produced under elicitation homogeneous (Shadrova, 2020; Lüdeling et al., 2021; Wan, 2021). In our case, the high school students prompted for Kobalt have all learned to produce the type of argumentative text prompted for our corpora in the very same classroom over years. While this does not mean they all attempted the same register or genre in production, it is plausible to assume that their acquisition background is very similar in that respect.

aspects, mode, genre, or register (Biber, 2012; Biber et al., 2016; Szmrecsanyi, 2019). More language-internal factors include stratified variance triggered by linguistic environments, such as the presence or absence of certain constructions of lexemes that may predict certain grammatical expressions, e.g., dative alternations or subject realization (Bernaisch et al., 2014; Deshors and Gries, 2016; Arroyo and Schulte, 2017; Cacoullos and Travis, 2019). Factors, whether they are language-internal or language-external, are mapped to predictable shifts in linguistic expression.

Needless to say, this perspective in the context of SLA research has fostered discussions around the necessity to redefine "the" native speaker, namely through "underscor[ing] the dangers of assuming what the target of L2 acquisition is" (Birdsong and Gertken, 2013, p.118). It has also raised attention to the question of how to carefully choose and specify what kind of group can legitimately serve as a control group for learner studies, for instance learners of other L2s, bilingual native speakers, instructors in a teaching setting, etc[4].

This specification of the composition of the control group does not, however, constitute a break with the broad paradigm of L1/L2 comparison based on frequency of occurrence of linguistic elements—a comparison that only makes sense if a certain stability can be expected within a group or environment. The work we present here takes a closer look at differences that go beyond unanimous, clearly external factor-dependent shifts in a data set across speakers, highlighting linguistic expression at the level of individual text production and the challenges of its quantification. This is relevant since all text—whether we find it in large-scale, general corpora or in smaller, task-based corpora is the result of individual text production.

In the following sections, we will first give a short and necessarily broad introduction to the theoretical framework of some strands of usage-based linguistics as far as they concern language learning. We then briefly discuss previous research of individual differences in corpus linguistics and present the two corpora used in this study. Following this, we discuss our results and look into the role of priming as a possible explanatory concept for higher degrees of variation with the aim of highlighting the relevance of both inter- and intraindividual variation in L1. In the final section, we summarize the conclusions we draw from our observations for a) corpus methodology in general and b) theoretical aspects of usage-based linguistic frameworks in particular.

2. LITERATURE REVIEW

In this section, we will briefly introduce theoretical models as they touch aspects of our data analysis, and review previous literature into L1 variability in learner corpus research as well as priming as a procedural factor in language production. We will summarize the main points as they relate to our research question at the end of the section.

2.1. Native Speakers and the Concept of the Target Language

It is well-known that linguistic theory has long since been divided into more rationalist, Universal Grammar (UG)-based approaches vs. more empiricist, behaviorist approaches subsumed under the umbrella term of usage-based linguistics. This has abundant implications for the explanatory models including questions of learnability, the role of frequency (if any), the status of target language vs. native language, the relevance of input in language acquisition, as well as study design and the operationalization of concepts in both paradigms. We approach our data from a usage-based framework and will hence not discuss UG-based approaches here[5].

As Ortega (2015b) points out, usage-based approaches do not constitute a single monolithic framework, but describe a habitus in the Bourdieuian sense, i.e., a set of socially learned and constructed ways to perspectivize language that challenge the previous status quo in many subfields. Central assumptions guiding the methodology and theoretical embedding are summarized in Larsen-Freeman (2006), Ellis and Wulff (2015), and Ortega (2015a). All usage-based approaches share the goal of describing and explaining linguistic patterns from observable language as it occurs in corpus or experimental data directly. Grammatical phenomena are mainly modeled in a variety of constructionist approaches, such as various strands of construction grammar (Goldberg, 1995, 2006; Croft, 2001; Sag, 2012; Boas, 2013), which are tightly intertwined with emergentist approaches to learning. Other approaches are shaped through socio-linguistic, variationist (for an overview see Geeslin and Long, 2014), and ethnographic perspectives.

Relevantly, the word *usage* can take on different scopes in different approaches and even within a single framework. Most generally, usage-based linguistics takes a behaviorist and empiricist view on language in that it seeks to describe linguistic behavior as it occurs. In modeling language acquisition, it takes the stance that language is also learned from and through usage (in emergentist/connectionist approaches). However, what constitutes usage can still differ even within this paradigm. For example, usage can be described in terms of concrete linguistic realizations (for instance by how much inflectional morphology is used) or in terms of the interactional, dialogical content of what two or more speakers experience in usage. In our research, we focus on the concrete linguistic realizations, because we have access to them more or less directly through the writing of our participants, and because we find it helpful to first document the linguistic reality as we find it in corpora, before we connect it to language-external factors.

In connectionist/emergentist models, L1-like competence is modeled as the result of a construction process using language input to arrive at linguistic abstractions and entrenchment

[4]The desirability of the latter is relativized by Birdsong and Gertken (2013, p.118) insofar, as "knowing more about the nature of natives' linguistic system is of theoretical significance in its own right".

[5]Generative grammar perspectives on individual differences, the role of input, and ultimate attainment can be found in Cook (1991), Borer (1996), Cook (1991), Hilles (1991), White and Genesee (1996), Yang (2004), Rothman and Iverson (2008), Rothman and Iverson (2008), and White (2015).

of auditory signals as well as abstract signs (Ellis, 1996; Bybee, 2002; Hoey, 2005; Tomasello, 2009). Importantly, construction grammar traditionally poses a unified space for all types of constructions from words through morphological units to syntax, famously summarized in Goldberg's "it's constructions all the way down" (Goldberg, 2006, 18) and playfully exaggerated by Boogaart et al. (2014, 1) as "it's constructions *all the way everywhere*." L1- and L2-learning across their linguistic (phonological, lexical, morphological, syntactic, semantic, pragmatic) levels are all equally attributed to frequency-leveraged mechanisms, and ultimate attainment in L2, including any of its limitations, is conceptualized as a function of input and usage. The native and the learner's target language systems do not differ in their underlying general quality, but in the input-dependent entrenchment of words, collocations, categories, and constructions. These are subject to constant change in both the native speaker and the learner and can be observed and analyzed in language output, i.e., experimental and corpus data[6]. Consequently, as Ortega (2013) points out, the distinction between learners and native speakers, that for a long time has so consistently been drawn even in studies dedicated to usage, becomes less and less relevant. This is also exemplified by some approaches in the area of language contact research (Backus, 2021).

The categorization and idealization of the native speaker in some of linguistic theory has been further deconstructed from a socio-historical (Bonfiglio, 2010) and sociolinguistic perspective (for instance, in a range of contributions to Doerr, 2009). Equally, the concept of nativelikeness in SLA research has been problematized from a language variability perspective by Birdsong and Gertken (2013) and others. These discussions are fueled by what has been described as a turn toward bi- or multilingualism in SLA research (Ortega, 2013; Geeslin and Long, 2014)—including the realization that multilingualism is, and has always been, the norm in language acquisition; that standardization of language is a fairly recent and often politically guided process; and that, while a speaker's language output in their various languages can be studied separately, their language system(s?) effectively cannot.

[6]There is of course more to say about the similarities and differences between usage-based and nativist language acquisition theories. One could argue that the introduction of the concept of *learned attention*—blocking of certain categories, constructions, etc. by means of the learner's L1 for acquisition of L2-categories and constructions—is rather close to the idea of an innate principles and parameters already being set (and thus, in a similar way, "occupied") by the L1, as in generativist approaches. Similarly, the covertness of entrenchments, and the limitations to their access, resembles the hiddenness of UG's *competence*. The predictions of usage-based accounts about what a learner opposite a native speaker knows can still be different from generativist accounts, even given the same data. Since in a usage-based account performance and its distribution is seen as direct expression of the underlying entrenchment, the same performance by a learner and a native speaker would result in assuming that their categories, words, and constructions are entrenched in a similar way. A generativist account first of all might not consider the data valid for taking hold of the underlying competence in either of the two, and secondly, might not conclude that similarity in performance means similarity in competence.

2.2. Previous Research Into Inter-individual Differences

Individual differences between speakers have raised attention in SLA research as factors determining the trajectory, velocity, and success of the learning process, as well as performance as a function of skill and other determining factors. Some of the observations pertain to language-internal or language-specific factors, such as shape of context, profile of the material that has already been uttered, and that is being planned (Szmrescanyi, 2006; Jaeger and Snider, 2013); language situation (Wiese, 2020); vocabulary (Kidd, 2012); or attainment as measured in production or reception/acceptability judgment (Dąbrowska, 2018; Birdsong, 2021). Much research has considered cognitive factors (e.g., aptitude, including as a function of age, cf. Berman and Nir-Sagiv, 2007), working memory, executive function, statistical learning faculty, intelligence (Skehan, 1989; Bates et al., 1995; Dörnyei, 2005; Kidd, 2012; Kidd et al., 2018) as well as more general psychological factors (attention, see, e.g., Roelofs, 2008; motivation, Lowie and Verspoor, 2019). Some consider external influences on language performance (time limit, test mode, channel, see, e.g., Ruth and Murphy, 1988; Chapman, 2016). Kidd et al. (2018) argue that individual differences result from a complex interplay of systemic cognitive and environmental factors and warn against downplaying variance in learner data as error variance, if individual differences are poorly taken into account.

In principle, all of these factors could also influence native speakers. However, where individual differences in L1 have been considered, this has mainly been done from a psycholinguistic perspective, e.g., Mulder and Hulstijn (2011), Dąbrowska (2012), and Birdsong and Gertken (2013). As a requirement for the direct comparison of L1 and L2 data, it is necessary to also gain an understanding of expectable differences among the L1 group. But L1 variability has only begun to gain awareness in L1/L2 comparison studies. For example, Mulder and Hulstijn Mulder and Hulstijn (2011) call for taking into account variability between native speakers in future SLA research, but still do this from a stratified perspective (by age; level of education). Similarly, Birdsong and Gertken (2013) discuss the necessity for a differentiation of groups in L1/L2 comparisons by consideration of inter-individual differences within and across groups. They argue (and we agree) that comparing the two groups can still be considered a legitimate method in SLA research as long as it is based on a differentiated analysis.

2.3. The Contrastive Paradigm

In spite of the theoretical possibilities provided by usage-oriented frameworks, variability in learner data has usually been investigated with contrastive/comparative methods e.g., *Contrastive Interlanguage Analysis* (CIA) in which a presumably homogeneous control group of native speakers is used as reference (Granger, 2002, 2015; Ädel, 2015). Effectively, even in these approaches that are sensitive to inter-individual variation, intra- and inter-individual differences in L2 data are used as indicators of the level of target language competence (Ädel, 2015;

Gablasova et al., 2017). Frequencies are modeled as dependent variables or expressions of underlying characteristics, such as target language competence. This implies that frequencies in target language are distributed within predictable and stable ranges, i.e., stationary. If frequencies were not stationary in the target language, but showed high variation, an approximation to target language frequency ranges would not be possible to achieve because the target of the approximation itself would be moving[7].

If frequency is expected to be stable and approximation to L1-like distributions is modeled as indicative of target language competence, this raises questions with respect to the adequate object of comparison. Obviously, learners cannot be expected to produce frequency distributions as they are common in newspaper or general-purpose corpora, but that is not necessarily due to lack of target language competence. Rather, newspaper or general-purpose corpora do not represent the speech of a single speaker, but are thematically and stylistically variable collections of text that are not representative of any one speaker of a language (Biber, 1993). A better object of comparison would thus be provided by the learners' input, for example through text books, assumed speech environment, and instructor speech. For example, Linford et al. (2016) investigate subject realization based on how much the assumed input and the output of their learners of Spanish match or diverge. To that end, they take a local corpus of native speakers formed under the same circumstances as their non-native-speaker corpus, and compare subject realization depending on the verb it occurs with, its frequency, and switch reference. They then examine the same measures on what they call a global corpus, Davies (2002)'s oral part of the *Corpus del español*. Indeed they find that the choice of corpus for comparison yields divergent results, i.e., that a certain distribution in the assumed input would lead to the conclusion that non-native speakers reproduce their input, while another distribution in another sample stipulated as input would lead to the conclusion that they do not, or to a different extent[8]. Although this research crucially depends on the recognition of situational variation and advocates the use of specialized corpora, it still does not consider the possibility of interference from inter- or intra-individual variation among native speakers.

In conclusion, despite the fact that many of the variables in the literature around individual differences are by no means specific to learners (Granger et al., 2015), and even though recent work has shifted the conceptualization of native speakers away from being a monolithic group, even studies that consider variation do not do so on an inter-individual level in L1 groups used for contrastive comparison.

Obviously, any speaker group characterization unavoidably carries some loss of information, since reductionist categorization implies the abstraction away from an object of study (Hulstijn, 2015). This is also the case with the group of native speakers, where, in addition to the information loss through categorization, a form of idealization tends to facilitate the assumption of homogeneity (Doerr, 2009; Davies, 2011). This may not be overall justified, as for example Dąbrowska (2012) shows considerable individual differences between native speakers of English in terms of inflectional morphology, passives, quantifiers and complex subordinating clauses. This poses challenges to the widespread idea of a definable subset of shared grammar between native speakers, which is a fundamental assumption in different theoretical strands of SLA research. Dąbrowska (2012) and DeKeyser (2012), as well as Birdsong and Gertken (2013) criticize the negligence of this fact, especially given that these differences cannot (only) be attributed to sociolinguistic factors. Birdsong and Gertken (ibid.), aside from questioning the overall comparability of monolingual native speakers with bi- or multilingual non-natives, call for careful methodological consideration of this. In the same manner, Hulstijn (2019) notes that the claim of great differences between adult native speakers serving as control groups in SLA research is still lacking a robust empirical underpinning. Our aim is to address this need for research and to illustrate native speaker variability from a corpus linguistic perspective from a group that would be predicted to behave homogeneously, following the literature.

2.4. L1 Variability in Learner Corpus Research

For corpus linguistics, Gries and Deshors (2014) diagnose a research deficit with respect to differences between native speakers which are used as a reference for learner language. Their analysis of the use of the modal verbs *may* and *can* in English L2 and L1 demonstrates variability among both groups. This is done with multifactorial regressions involving interactions between fifteen different factors like syntactic characteristics of the clause and various morphological and semantic features of the subject. They use a method entitled Multifactorial Prediction and Deviation Analysis with Regressions (MuPDAR), which shows statistical interactions of lexical and syntactic elements in large corpus data[9]. While the authors themselves describe this method of analysis as very complex and challenging, our work will illustrate that inter-speaker variability in L1 data can also be examined and demonstrated with less demanding analytical methods, and with smaller, more controlled and deeply

[7]Some studies that explicitly take into account variance in the input that learners are exposed to do so on the basis of group characteristics. Conceptually, these studies model differences between the *input* for learners and non-native speakers. This is unlike the central question of early SLA research, that often had a more deficit-oriented perspective, and asked to what extent the *output* of learners conforms to that of native speakers (for a critical review, see for instance Klein, 1998). A method to determine both is the comparison between corpora. Notably, even despite not explicitly stating any adherence to Contrastive Interlanguage Analysis (CIA), studies under a variationist umbrella that use naturalistic data often make very similar methodological decisions in comparing their corpora. In fact, Eskildsen and Cadierno (2015) describe the different foci of linguistic patterns vs. sociolinguistic classifications as cognitive/usage-based (CUB-SLA) theories on the one hand, and theories based on conversation analysis (CA-SLA) on the other, and view them as complementary rather than mutually exclusive.

[8]A problematic aspect of this study is the definition of "frequent" vs. "infrequent", which is pragmatically—and understandably—drawn arbitrarily at the 1% threshold of verb tokens in the corpus. However, this will necessarily massively fluctuate with corpus size and type. Where theoretical conclusions from frequency are drawn, the set of affected words should be stable, but derived in this way, it cannot be stable.

[9]For critical stances toward lexical statistics (see Shadrova, ress; Kilgarriff, 2005; Schmid, 2010; Koplenig, 2017).

annotated corpus data. This offers a more widely accessible approach to comparative SLA corpus studies, and, since smaller data can be manually annotated, allows for the analysis of a greater variety of linguistic phenomena (cf. Lüdeling et al., 2021).

According to Granger (2002), native speaker corpora provide relevant information on the frequency and use of words, phrases and structures. Occurrences and co-occurrences of certain linguistic features can be used as a basis for comparison between L1 and L2, concretely of L2 mis-, over-, or underuse (Granger, 2002; Ädel, 2015; Gablasova et al., 2017). Frequencies in L1 serve as a benchmark for the frequencies of the same features in learner language and thus play a central role in comparative methods such as the CIA in SLA research. This is a consequence of the idea of entrenchment as a direct neuronal correlate of frequency in the input. Divjak and Caldwell-Harris (2015) in a literature review present the discovery of characteristics that correlate with frequency (e.g., word length, concreteness, age-of-acquisition of a word/structure) as well as the evolution of contextualized frequency measures, such as dispersion (homogeneity of the distribution of a word in a corpus) or surprisal (how unexpected a word or sequence is, given its context). Since Langacker's (1987) introduction of the concept of usage-based learning, there has been continuing research for "the measure which is best suited to predict entrenchment" (Divjak and Caldwell-Harris, 2015, p.67). This concerns, among other things, the granularity level at which frequencies are measured along with the question of the units that are effectively entrenched (for example words, morphosyntactic categories, phonetic sequences etc., cf. for example Ellis, 1996; Croft, 2001; Bybee, 2002; Wray, 2002; Goldberg et al., 2004; Bybee and Torres Cacoullos, 2009; Ellis and Frey, 2009).

One of the few studies to our knowledge that deal with the challenges of native speaker variability in the frequency of occurrence of linguistic structures is Gablasova et al. (2017) investigation of four linguistic features in five L1 corpora of informal spoken English: a concrete co-occurrence (*I think*) and word form co-occurrences (adverb+adjective), as well as past tense and passive occurrences. They emphasize the necessity of investigating inter-speaker variation within corpora before comparing frequencies across them, because they consider it equally important to reflect on possible causes of variation between corpora, which could, for example, be due to different corpus designs, subject groups and data collection methods. The results illustrate that corpora of similar native speaker language can differ remarkably, both within and across corpora.

2.5. Priming and Corpus Data

So far, we have introduced relatively stable or situational factors that may lead to inter-individual differences. Those are either non-linguistic (age, region, gender); language-related (aptitude, reading experience); or fully linguistic (lexical and syntactic environment). Those affect the linguistic behavior of a speaker in generalized ways across their production (although some of them may still fluctuate over time). Another factor that affects language production is priming, i.e., the semi-persistent activation of elements that facilitates their repetition or the co-activation of other elements based on similarity of structure or content. For the purpose of this paper, priming can be understood as a mechanism that temporarily raises the probability of a word or category to re-occur after it has been introduced.

Priming or persistence started getting attention from a corpus-linguistic angle only during the past 15 years. It is at the intersection of cognition and factors inherent to the linguistic system. Its psycholinguistic underpinnings and exact mechanics are not fully understood, but the linguistic dimensions of its occurrence, as well as conditions that favor it, have been given some attention in the literature (for an overview, see Gries and Kootstra, 2017).

Priming can occur as a particular form or as a pattern (Szmrecsanyi, 2005; Szmrescanyi, 2006; Gries and Kootstra, 2017), or, as we understand it for our purposes, as lexical or structural priming, for example a morphological class rather than a specific word. If priming had an effect on the morphological level in our data, a morphological class once introduced would re-occur at higher rates than if it had not been evoked, in effect forming clusters in a text. Speakers are susceptible to other-priming (priming by external factors, such as the prompt or interlocutor speech) as well as self-priming by their own text-production. Since priming is a procedural phenomenon, its effects decrease with a higher prime-target distance. This means that it may affect *only part* of a text, making it very different from more stable factors, like age or reading experience, or even the more fluctuating, like motivation, which will still affect the *whole* text that a participant contributes. This is relevant to the methodological and theoretical model because it highlights the fact that cumulative corpus counts are not a single, but a twofold dimensionality reduction that collapses both the inter- and intra-individual variability that exists in a corpus, i.e., two ranges, into a single number.

Gries and Kootstra (2017) suggest that corpus linguistic studies are suitable for exploring priming effects, as they provide a more natural usage-based perspective on priming than psycholinguistic experiments with potentially unnatural stimuli. This specifically affects prompt-based and self-priming. Chapman (2016, p. 110) in a study of second language writing assessment shows that lexical sophistication, academic vocabulary use, syntactic complexity, cohesion, and fluency of a response can be strongly influenced by prompt characteristics. Even relatively abstract elements such as the morphological class of particle verbs in German can be prompt-primed in both L1 and L2 according to Lüdeling et al. (2017). The way writers respond to a specific prompt is also expected to have more far-reaching consequences, namely on the selected register of the produced text[10]. Although priming exists on all linguistic levels (phonetic/phonological, semantic, pragmatic, syntactic, discoursive, etc.), we will only consider structural morphological priming, which we will discuss in section 4.4.

2.6. Research Question

The research question guiding our analysis can be summarized as "how variable are German native speakers from a highly homogeneous group in their distribution of a) morphological

[10]This has also been discussed for one of the corpora used in this study, Kobalt, in Shadrova (2020, ch. 7).

subclasses of nouns and verbs; and b) higher-order syntactic elements in task-specific, highly controlled corpus data?", or, simpler put "what kind of information with respect to inter- and intra-individual variation would we lose in the cumulative analysis of our corpora?"

We enter from a learner corpus-oriented research paradigm, but we will not look into learner data in this study—instead, the observations we report are born from *intended* comparisons with learners within a connectivist and emergentist usage-based framework.

3. MATERIALS AND METHODS

The texts used in our study were written by participants of the native speaker control group in the collection of the two German learner corpora Kobalt (Zinsmeister et al., 2012) and Falko (Reznicek et al., 2012). Both corpora are comprised of prompted argumentative essays written under controlled conditions (90 min, handwritten or typed without aids such as dictionaries). Kobalt contains 20 L1 texts, in Falko we use 95 L1 texts for the morphological analysis and 65 for the syntactic categories. We are forced to accept this limitation since not all L1 texts in Falko are available with corrected dependency tags yet. Neither of the corpora was compiled for the purpose of this study, both are publicly available (see data availability statement at the end of this paper) and have been previously used in a number of other studies (Hirschmann et al., 2013; Zeldes, 2013; Hirschmann, 2015; Lüdeling et al., 2017, 2021; Shadrova, 2020; Wan, 2021, among others).

L1 contributors to both corpora were chosen from a very homogeneous group, 12th year high school students from the same school in Berlin in the Kobalt subcorpus, and early college students from Berlin as well as high school students from Berlin and Potsdam (a smaller city near Berlin) in the Falko subcorpus. This way, we were able to control for age, region, urban vs. rural influences, and even exposure to the same teaching materials in the case of high school students. We did not control for socio-economic status directly, although both high schools were chosen from more affluent parts of town for practical reasons. Unfortunately, the reality of the German education system is highly selective and stratified. We do not expect that there would not be any differences at all between our participants or their parents with respect their socio-economic status or education background. However, based on German population statistics, we can assume a high level of homogeneity based on the group selection and the social reality in Germany[11].

Both corpora are prompt-based and controlled with respect to topic. In Kobalt, the prompt is *Geht es der Jugend heute besser als früheren Generationen?* "Do young people today do better/have a better life than previous generations?" In Falko, participants were free to choose from four different prompts on topics attempting to elicit a discussion of controversial points of view. The topics that were chosen for corpus collection resemble the ones used in the ICLE corpus, cf. Granger et al. (2020).

- *Kriminalität zahlt sich nicht aus.* ("Crime does not pay off", labeled *crime*);
- *Die meisten Universitätsabschlüsse bereiten die Studenten nicht auf die wirkliche Welt vor. Sie sind deswegen von geringem Wert.* ("Most university degrees do not prepare students for the real world. They thus are of low value," labeled *university*);
- *Die finanzielle Entlohnung eines Menschen sollte dem Beitrag entsprechen, den er/sie für die Gesellschaft geleistet hat.* ("A person's financial remuneration should depend on the contribution that they make to society," labeled *incentive wage*);
- *Der Feminismus hat den Frauen mehr geschadet als genutzt.* ("Feminism has done more harm than good to women", labeled *feminism*).

Neither elicitation was based on school work or homework or graded in any way. Participants contributed texts of variable length. In Falko, text lengths range from 181 to 1728 tokens including fluctuations by topic (min. 217, 284, 181, 436; max. 1728, 1305, 1335, 1184 tokens for the topics crime, feminism, incentive wage, and university respectively; mean: 822.20, 886.46, 872.17, 871.88; median: 712, 915, 846, 978). In Kobalt, text lengths range between 483 and 813 tokens (mean: 624.45, median: 644.5).

Both corpora contain metadata on the participants' linguistic background (language biography, i.e., L1s and L2s with age at the onset of acquisition, years of training, years of immersive exposure). These were identically collected in the L2 subcorpora of both corpus projects, but are highly uniform in our L1 subcorpora, with barely any early bilingual speakers and no longer interruptions of L1 immersion. Kobalt additionally contains scores from a standardized c-test (onDaF, now onSET, Eckes, 2010). We did not find correlations between the frequency of morphological forms including a binary distinction between complex vs. simplex forms on the one hand and gender or high school vs. college students (i.e., level of education, self-selected group of language students) on the other. No other correlations were found with other aspects of the available metadata either. We will hence not address this issue further.

3.1. Methods

We present descriptive statistics, using relative frequencies (normalized to all occurrences of verbs in Kobalt and nouns in Falko) and proportions of categories normalized to 100%.

[11] After primary school (year 4 or 6, depending on the federal state), students are divided into three general tiers, so-called *Haupt-* and *Realschule* and *Gymnasium*. *Haupt-* and *Realschule* end after year 10 and aim to prepare students for vocational training, which is accompanied by ongoing education at professional school *Berufsschule*. Students completing their studies with the high school degree *Abitur* at *Gymnasium* acquire the right to study at a university or college. The separation of students into tiers attracts much critical debate for being known to be a highly socially selective procedure restricting upward mobility. Some schools offer integrated schooling (*Gesamtschule*), but they still follow the principle of separate degrees, and students are usually taught separately by attempted degree in several subjects. Options to enter university without *Abitur* are very limited, especially outside of medical or engineering subjects, from which we did not collect data. In Germany, only 21% of the children of parents without academic degrees begin college studies, while 74% of parents with academic degrees do. Ratios are even more contrasted in the humanities, and also locally, since universities use cut-off marks based on student's *Abitur* grades to limit admissions, which affects Berlin in particular. *Abitur grades* are further known to correlate with socio-economic status to a lamentable degree.

We computed regressions for potential text length dependency, since text length is well-known to correlate with many corpus linguistic measures. In our data, text length correlates highly with simplex verbs and nouns, but not with any of the other categories (see section 6.1 in the **Appendix**). We include plots of randomized samples of the original lengths in the appendix to show the expected variance if categories were randomly distributed, confirming our conclusion that text length is, somewhat surprisingly, not a meaningful factor in morphological category distribution.

With the exception of regressions for text length, we limit our statistics to basic descriptive measures such as percentages and simple variance computations, since we are mainly interested in the composition of categories from subclasses. Accounting for the variance of several factors in a system in a single measure necessarily involves a dimensionality reduction that we are not ready to perform on this data, because we have limited understanding of its linguistic repercussions. In addition, from the results we obtain in the comparison between native speakers, we cannot be sure that frequencies converge. This limits our trust in the abstractability of relative frequencies from this data to idealized probabilities—we are not confident in that the data is ergodic and stationary (Shadrova, ress; Piantadosi, 2014; Dębowski, 2018), or can truly be seen as a random sample from a population in the statistical sense. If it were not, the central limit theorem would be caused to fail and inferential statistics would be rendered undefined. More clarification of the mathematical underpinnings of those categories as they occur in corpora are required before we can proceed with inferential statistical modeling, such as regression. This remains for future research.

We further present a sliding window analysis for a discussion of priming as a factor that could potentially contribute to high variability. For this, we have defined overlapping windows of 50 tokens each, the first covering tokens 1-50, the second 2-51, the third 3-52, and so on. Each text is represented by $textlength - 49$ windows. Data points show cumulative counts of the occurrence of the respective category in each window. Colors differentiate between the total token occurrence of the category and the number of different lexemes (types). For example, a category can be represented in 5 tokens and 3 types within 50 tokens, i.e., one type would be repeated three times, or two would be repeated twice in that window. For most of the windows, the number of types equals the number of tokens. Window size was chosen arbitrarily, but attempting to maximize representation of peaks and slumps. If window size is chosen too large, two peaks might be bridged, making it appear as though the category was uniformly represented across the whole window. If window size is chosen too small, accumulations are not properly represented. A better understanding of correct choice of window size should be derived from future research in alignment with psycholinguistic observations.

All analyses were performed using R (R Core Team, 2015) on RStudio (RStudio Team, 2015) with packages dplyr (Wickham et al., 2018), reshape2 (Wickham, 2007), and ggplot2 (Wickham, 2016).

3.2. Annotations and Categorization

We investigate structural variation on several levels of complexity and abstraction as we expect that the amount of linguistic material involved in a structure may influence the range of variability. As representatives of a higher level of interdependent structure, we examine syntactic dependencies and part-of-speech distributions. For more fine-grained categories, we look at the morphological and morphosemantic subclasses of nouns and verbs.

Both corpora are part-of-speech-tagged with TreeTagger (Schmid, 1994) and the Stuttgart-Tübingen tagset (Schiller et al., 1995), and dependency-parsed with manual correction of dependencies (MaltParser, Nivre et al., 2006 with Foth, 2006's dependency grammar). The part-of-speech tagging and dependency parsing are generated on the target hypothesis, a normalization layer that consists of a hypothetical reconstruction of an orthographically and syntactically correct version of the text (Reznicek et al., 2013). Lexical items are not corrected or changed except for orthography. This method was designed for L2 data, but even for essays written by L1 speakers, automatic parsing does not yield satisfying results when based on the original document, hence the need for a normalization layer.

Morphological categorizations of nouns (Falko) and verb-type classifications in terms of syntactic category and morphosemantic components (Kobalt) were manually annotated. Detailed annotation schemes for both classifications can be found in the **Appendix** in **Supplementary Materials**[12].

Nouns in Falko were classified according to the word formation processes underlying their structure, for example as determinative compounds, derivations, nominalizations, etc. The annotation followed the guidelines in Lukassek et al. (2021) that were developed in several iterations of test annotations by two or more annotators, discussions of the results and refinements. Guidelines were furthermore tested by three independent annotators whose inter-annotator-agreement (Artstein and Poesio, 2008) for the annotation layer reported in this paper was perfect (Fleiss' $\kappa = 0.81$).

Lexical verbs in Kobalt were classified with respect to their morphosemantic properties (simplex vs. complex, i.e., particle or prefix, vs. support verbs). More detailed information on these classes will be provided in the next section. Syntactic verbs were classified according to the syntactic environment they trigger (modal, modifying, auxiliary, copula, constructional verbs). Simplex, particle, prefix, modal, modifying, auxiliary, and copula verbs are easy to classify because they occur in very clearly defined syntactic environments or have a distinct shape (prefix, particle, simplex verbs). Support verbs and constructional verbs are subject to more ambiguity, since they mark a deviation from the semantic or syntactic norm. More detailed information on these annotations can be found in Shadrova (2020, section 3.2)

[12]Since manual annotation is laborious and resource-intensive, we refrain from adding the complementary annotation layers to the respective other corpus, although, obviously, nouns were also used in Kobalt and verbs in Falko. Since we do not attempt a direct comparison between the two corpora, this should not constitute a problem.

and in a Zenodo repository which also contains the annotated data: 10.5281/zenodo.3584091.

4. RESULTS

4.1. Kobalt: Verb Subclasses

We first investigate the distribution of subclasses of verbs in Kobalt. For this, we will look into morphologically and syntactically defined subclasses. For the syntactic subclasses, we consider auxiliaries, copula verbs, modal, and modifying verbs, as well as verbs in constructional use (see **Appendix** in **Supplementary Materials** for annotation guidelines)[13]. Morphologically complex verbs in German include prefix verbs that contain an inseparable prefix to a base such as *verlegen* ("misplace" vs. the simplex *legen* "to put, to place") and particle verbs, that include a separable particle to a base. The particle is split from the base in inflected verb forms, i.e., in non-analytical constructions (constructions lacking an auxiliary or modal verb), and forms a different participle. In the case of the particle verb *vorlesen* "to read out loud, to read to someone" vs. the prefix verb *verlegen* "to misplace," this occurs in following way: *Sie liest den Kindern die Geschichte **vor**; Sie hat den Kindern die Geschichte vor**ge**lesen* "she is reading/has read the story to the children" vs. *Er **verlegt** oft seine Brille; Er hat seine Brille **verlegt** (not: **vergelegt**)* "he has misplaced/frequently misplaces his glasses"[14]. Semantically complex verbs here refer to the difference between simplex verbs on the one hand and support verbs in support verb constructions (*Funktionsverbgefüge*), which take on a non-compositional, non-literal meaning in lexicalized VP-NP combinations, on the other. Morphologically complex verbs can also be considered semantically more complex because they tend to semantically extend their bases[15].

In our analysis, we are interested in the *composition* of the verb class with respect to its syntactic subclasses, not simply in the relative frequency of each subclass—how much space does each subclass take relative to the other categories[16]?

As we would categorize the phenomenon according to our guidelines, we would find a distribution as visualized in **Figure 1**. From this result, we could derive conclusions for our hypothesis—for example, that auxiliaries, copula verbs, and modal verbs are equally frequent; and that simplex verbs are the most frequent category, followed by prefix and particle verbs—and we could bring that together with SLA theory to hypothesize how those distributions might diverge in learners.

Or could we? The boxplot in **Figure 2** accounts for the variance between documents in each category. Here we can see that there is in fact considerable variation within and overlap between categories. While on average, the previous description still holds true to a degree, it no longer covers all of the data. However, even from this perspective we can still model category frequency to fit with the idea of an idealized, albeit strongly probabilistic native speaker.

But since usage-based theory presumes frequencies to be meaningful and reasonably stable aspects of linguistic expression, this wide frequency range raises our suspicions - what is going on in the L1 data and how do we consider it methodologically? The composition of subclasses in each text, represented in **Figure 3**, provides a clearer picture of the vast variability in frequency realizations in what would theoretically be a homogeneous L1 corpus[17]. Rather than just using "more" or "fewer" complex forms, each individual participant in the native speaker group appears to follow their own *distribution* of classes—a type of information that is, to a degree, implicitly included in the boxplot in **Figure 2**, but becomes strikingly more obvious in the tiled pie charts. While some participants use prefix verbs more than any

[13]We exclude the category *gehen_cx* from our analysis. It labels the verb *gehen* in the constructional use of *Wie geht es dir?* 'how are you doing?' and is used at deceptively high rates since it is part of the prompt. We believe it cannot be considered well in the analysis of priming either, because it is highly salient in this context and is used with clear intention for text structuring purposes. We believe it should thus not be compared with the other categories.

[14]Some complex verbs in German are not analyzed in the same way by all speakers. This is especially the case for complex verbs that incorporate nouns, like *staubsaugen*, as opposed to the phrasal variant *Staub saugen* ('to vacuum', literally 'to suck dust'). Some speakers read it as a prefix verb, *er staubsaugt, ich habe gestaubsaugt* ('he is vacuuming', 'I have vacuumed', literally 'he dustsucked', 'I have dustsucked')), while others read it as a phrasal unit (*er saugt Staub, ich habe Staub gesaugt*, literally 'he sucks dust', 'I have sucked dust'), which then lends itself to a particle verb analysis: *er saugt staub, ich habe staubgesaugt*. Similar patterns can be observed in newer verbs like *downloaden* ('to download'), *um das Video downzuloaden* (particle) 'in order to download the video' vs. *um das Video zu downloaden* (prefix). In Kobalt, there were only very few cases of this type, and they were analyzed as closely as possible to the original writing. If a participant wrote them as phrasal units, they would be analyzed as a simplex or a support verb, depending on the compositionality of the combination. If a participant wrote them as one word, they were analyzed as prefix verbs, unless there was syntactic evidence for the separability of a particle, as in *downzuloaden*. Often, speakers avoid commitment to one analysis by using syncretic forms, for instance *Man kann das Video downloaden, indem man auf den Link klickt* ('One may download the video by clicking the link', prefix or particle), rather than *Um das Video downzuloaden (particle)/zu downloaden (prefix), muss man auf den Link klicken* ('In order to download the video, one has to klick the link').

[15]In some cases the connection is synchronically fairly far removed, as in *raten* 'to guess, to advise' and *verraten* 'to betray'. A more detailed discussion can be found in a comparative study of complex verb productivity in German L1 vs. L2 in Lüdeling et al. (2017).

[16]This would matter in an actual contrastive scenario, because it can provide insights into the structure of the morphological system, for example whether morphologically complex verbs take an equally central position among the other subclasses in L2 and L1, and answer questions concerning productivity and lexicalization, complexity, or the development of aspect and perspectivation through verb modification. Lower morphological complexity as an L2 phenomenon has generally gained interest in the SLA literature in recent years (Zeldes, 2013; Ehret and Szmrecsanyi, 2016; Lüdeling et al., 2017; Yoon, 2017; Brezina and Pallotti, 2019; De Clercq and Housen, 2019, and others).

[17]We are aware that pie charts are not an ideal type of data visualization for most purposes, because they tend to make a comparison of exact proportions difficult. This is due to limitations of the human mind, that seems to be less well-equipped to compare and interpret dimensions from angles other than 90 degrees. However, in our case, we will compare a large number of compositions, i.e., distributions of several (more than three) factors. This becomes very difficult to read in stacked bar plots, since two or three factors can be ordered by relative size of each factor, but four cannot, making the bars and colors very noisy in perception. Pie charts have shown to be the most efficient at visualizing the differences both *between factors* and *between texts* in easily graspable ways. The central point is the distribution of subclasses within each category, i.e., whether the pies look similar or different in their division into pieces. If subclasses of categories were similarly distributed, all pies should be cut in similar ways, i.e., have pieces of similar shape and size, as is the case in some of the later plots in this paper. For the morphological subclasses, they tend to not be, and that is the point. Precise percentages or individual mappings do not matter much and will only be referred to for exemplification.

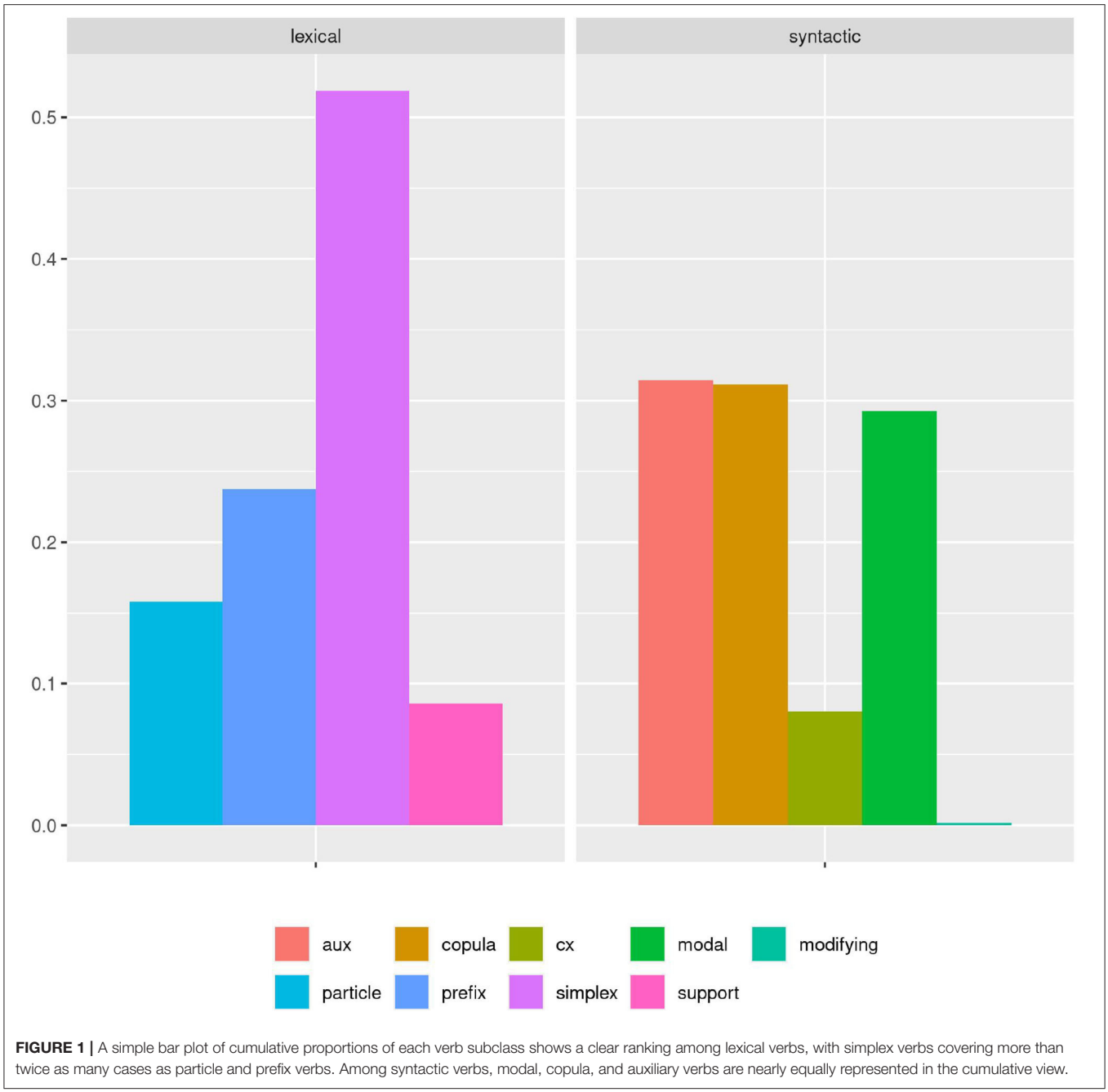

FIGURE 1 | A simple bar plot of cumulative proportions of each verb subclass shows a clear ranking among lexical verbs, with simplex verbs covering more than twice as many cases as particle and prefix verbs. Among syntactic verbs, modal, copula, and auxiliary verbs are nearly equally represented in the cumulative view.

other category (DEU_001, DEU_017), others use twice as many simplex verbs as all other types combined (DEU_005, DEU_011). Some use more support verb constructions than particle verbs (DEU_005, DEU_012, DEU_021), while for most others, support verb constructions make up the smallest part. Since quantitative corpus linguistics builds on the assumption that frequency of occurrence has meaning, this result is puzzling and slightly worrisome. Which one of those speakers should be considered representative of the target language for a learner?

Figure 4 shows similar diversity in the distributions of syntactic or functional verbs, i.e., verbs that occur in or trigger specific syntactic environments, such as auxiliaries, copula, or modal verbs. If speakers followed frequency distributions in their realization of words (by morphosyntactic category) or relational structures (to express modality or temporality), modals and auxiliaries should be distributed more equally. For instance, modal verbs are a) very schematic and transparent in their use and b) not very diverse[18]. Auxiliaries are even more limited

[18]German has six: *wollen, können, sollen, müssen, dürfen, mögen/möchten*, 'to want','to be able to', 'to be obligated; shall; epistemic must','deontic must, to have to', 'to be allowed to', 'would like to').

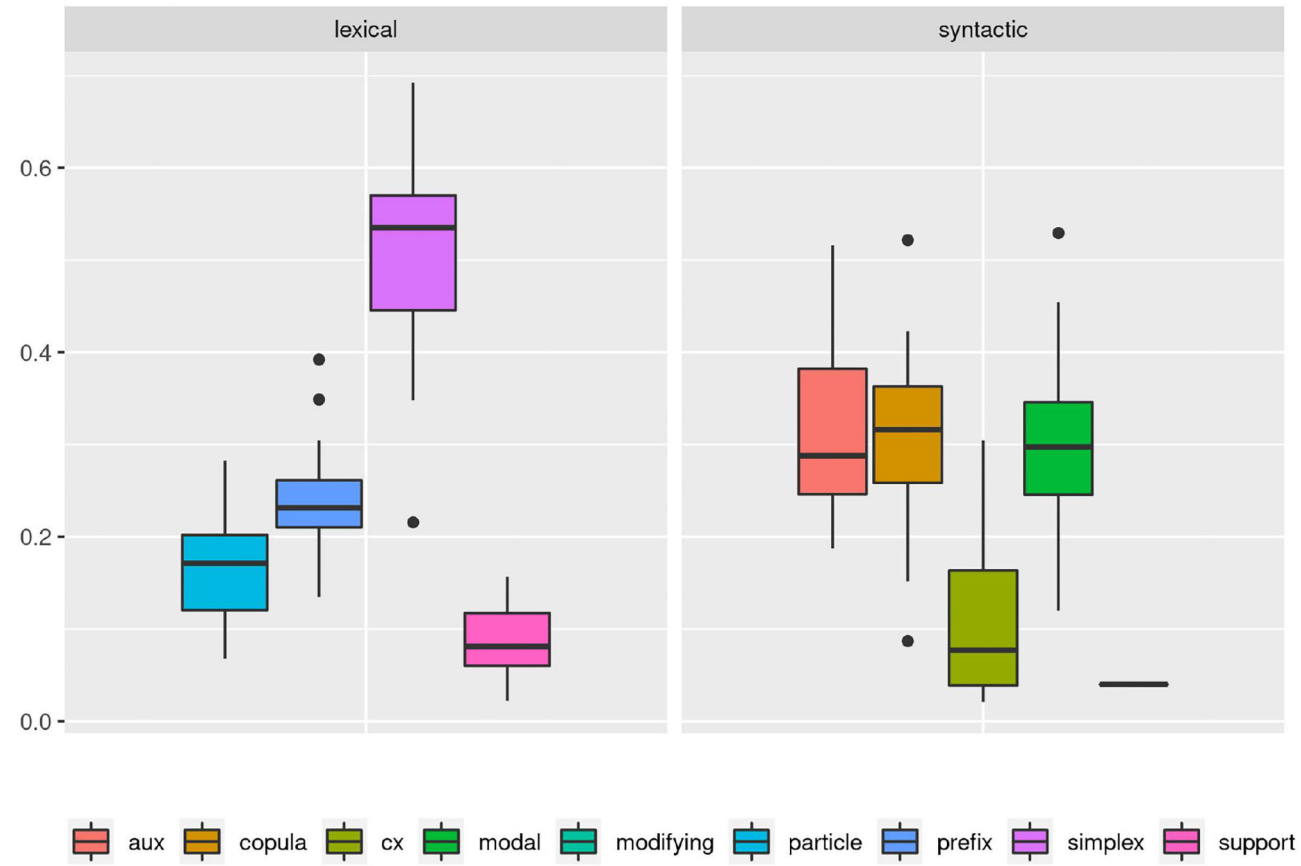

FIGURE 2 | A boxplot based on the proportion of each verb subclass as it is found in the individual documents of the corpus shows a more complex picture. Long whiskers indicate considerable overlap between all syntactic verb subclasses. Similarly, the non-simplex lexical verb subclasses all partially overlap and even the clear dominance of simplex verbs among the lexical verbs is called into question with several outliers overlapping between prefix and simplex verbs.

and equally transparent. However, in our data, the proportion of auxiliaries among syntactic verb forms lies anywhere between 19% (DEU_007) and 52% (DEU_018). Even more strikingly, the use of modal verbs among the morphosyntactic subclasses ranges between 0 (DEU_018) and 53% (DEU_005). If the same was found in a learner group, one might conclude that a learner avoids modal verbs due to incomplete attainment, but obviously, in a native speaker at high school level, this explanation is lacking.

4.2. Falko: Noun Morphology

In a similar fashion to **Figures 1**–**3**, **5**, **6** show the distribution of noun morphology in the Falko corpus. We first see a bar plot showing the cumulative distribution of morphological types of nouns across the corpus in **Figure 5** and then a box plot accounting for the variance between the 95 native speaker documents included in Falko in **Figure 6**. Both plots are divided by topic, because the topic may influence the chosen text type or register of the text, which in turn may trigger variability in linguistic realization. Obviously, in a text written in response to the prompt on *feminism*, we would expect a significant amount of nouns referring to adults of either female or male gender and to children. All of these concepts are realized as simplex nouns in German (*Frau* "woman", *Mann* "man", *Kind* "child"). Furthermore, the topic is introduced with a prompt, which in an analysis of morphological aspects of complex verbs in Falko has been shown to produce structural priming effects on the morphological level (Lüdeling et al., 2017)—both learners and native speakers use more particle verbs if the prompt includes a particle verb. At least for the university topic, the prompt yields a similar effect for nouns. The prompt features two non-native nouns, one of which is part of a compound. From **Figure 5**, we can see that *kdet* (determinative compounds) and *nnat* (non-native nouns) are the two most frequent classes, which we interpret in terms of a priming effect.

The document-wise distribution in **Figure 6** yields a more differentiated picture. Let us consider simplex nouns as an example. According to **Figure 5**, this noun type is prevalent in the *feminism* topic. However, in **Figure 6** we can see that individual texts include fewer simplex nouns than derivative nouns, which constitute only the fourth most frequent noun type in the cumulative distribution of the *feminism* topic. In the *incentive wage* topic, simplex nouns are one of the two most

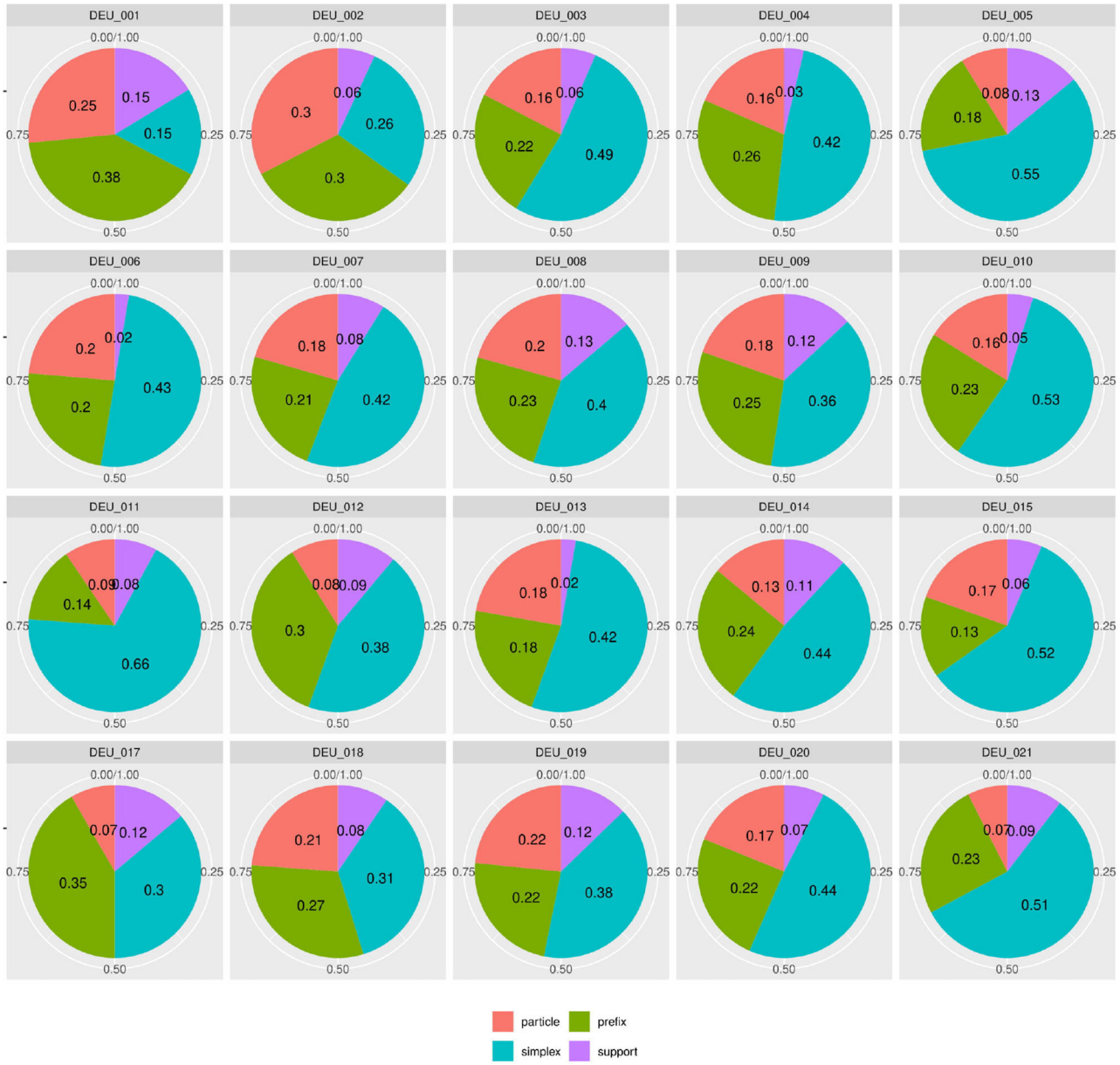

FIGURE 3 | Individual distributions of lexical verb subclasses in the native speaker documents in Kobalt show striking differences in proportions. Pie charts are usually not ideal tools for distribution visualization, but were in our view the best choice in this case, cf. footnote 18.

frequent noun types in the cumulative distribution. Nevertheless, we can see from the document-wise distribution in **Figure 6** that simplex nouns rarely occur in some of the texts.

Figure 7 shows document-wise distributions of each noun type. For better interpretability, we grouped the two concatenative word formation types compounding (*kdet*) and derivation (*der*) as well as the two non-concatenative types conversion (*kon*) and other nominalizations (*nom*). Due to space limitations, we only present selected distributions in **Figure 7**. Plots for the remaining texts can be found in a Zenodo repository (10.5281/zenodo.4752308). Within the texts from the *university* subcorpus, the differences for the concatenative class are most striking. Whereas in text fu082d_2007_10, concatenative word formation processes account for 36% of all nouns, in fu083d_2007_10, the concatenative group covers 59% of all noun occurrences. Similarly, non-native noun formation varies between 16 and 32% of all nouns in the respective texts (cf. fu080d_2007_10 vs. fu070d_2007_10).

A similar variance can be observed for the texts on *incentive wage*. As we can see from the document-wise distributions, the simplex nouns (*sim*) vary between 4% and 35% (cf. fitfu072d_2006_10 and dhw027_2007_06). Concatenative

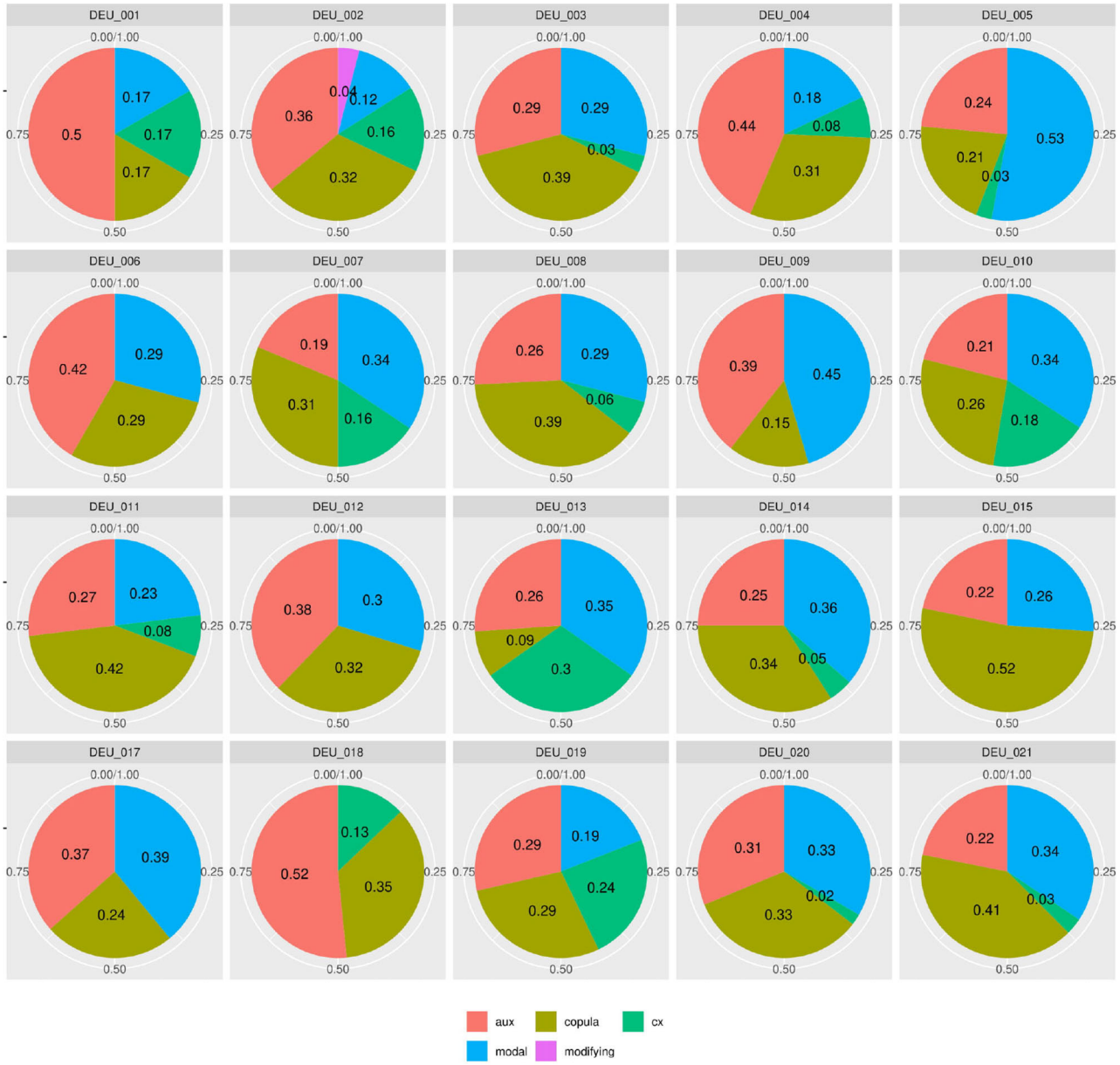

FIGURE 4 | Individual distributions of syntactic verb subclasses in the native speaker documents in Kobalt show striking differences in verb subclass proportions in spite of low degrees of lexical semantics present in the categorized verbs.

nouns (*der* and *kdet*) make up between 24% and 47% of all nouns (cf. dhw031_2007_06 and dhw030_2007_06). Non-concatenative nouns (*kon* and *nom*) vary between 16% and 37% (cf. dcs004_2007_10 and dhw031_2007_09). Between these extremes, varying sub-divisions of the noun spectrum are possible.

The strongest variance in distributions can be found in the subcorpus of texts on the *crime* prompt. Transpositions (*trans*) account for 2% to 31% of all nouns (cf. dhw026_2007_06 and dhw022_2006_06). Concatenative nouns (kdet and der) vary between 16% and 48% (cf. dew10_2007_09 and dew06_2007_09). Simplex nouns are being used between 6% and 42% of all occurrences (cf. dhw011_2007_06 and dhw010_2007_06).

In a nutshell, the distribution of morphological nouns in Falko shows that deriving insights about the frequency of noun classes from data accumulated over speakers is highly problematic. The fact that one class is prevalent in the overall distribution does not mean there cannot be individual texts with entirely different relative frequencies for the same class. This raises the question whether cumulated speaker data, at least for this phenomenon, is

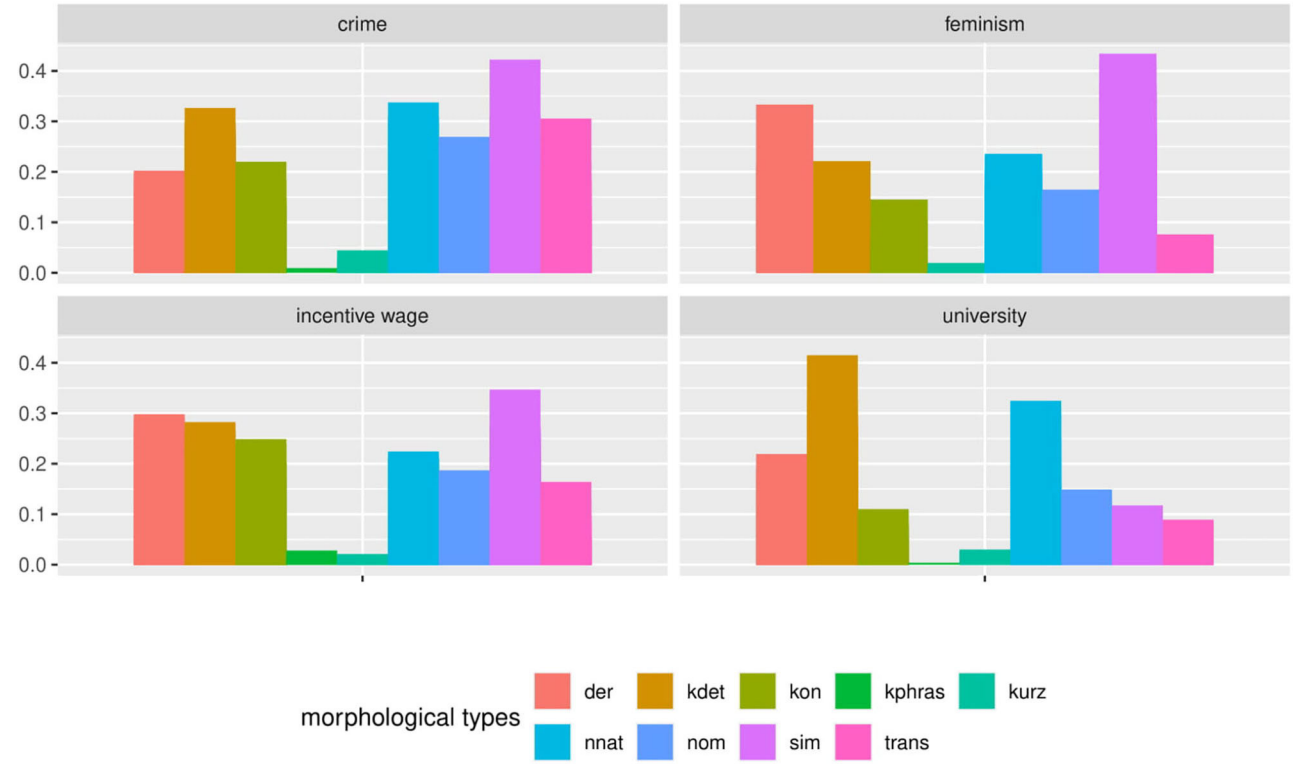

FIGURE 5 | The cumulative distribution of morphological types of nouns across Falko differs by topic. For instance, in the *feminism* topic simplex nouns are prevalent, whereas in the *university* topic non-native nouns and determinative compounds are the largest groups.

interpretable at all, or in other words, whether even situationally specified target language frequencies can be defined in the first place.

4.3. Syntactic Classifications Affecting the Larger System

However, such differences do not appear across syntactic categories. **Figures 8**, **9** show the distribution of parts of speech and syntactic dependencies in randomly selected texts from Falko and Kobalt[19]. Unlike the previous analyses, these plots show much more comparable realizations of category proportions. That is not to say that there is no variation at all—in fact, there is at least one text in the individual dependency distribution in Falko that sticks out with a much lower proportion of prepositional dependencies (dhw_010_2007_06, top row third from left) than any other text shown here. There is also some fluctuation in the proportions between the other types. However, overall, for most texts, distributions are roughly quartered between the four categories, or rather tend to be realized through attributes and object-type dependencies by about half, filling up the other half with 40/60 prepositional and other (verb and determiner) type dependencies. **Figure 8** shows that there are also some topic effects.

Similarly, parts of speech are distributed more equally between texts (**Figure 9**). This is not surprising, trivially following from **Figure 10** because dependencies are derived from parts of speech, but also because part-of-speech distributions are known to be language-specific with such clarity that they can be used for determining the native language of competent L2 speakers writing in their second language, and even the original language of a professionally translated text (Teich, 2003). In fact, the success of statistically based language parsing and translation is based on the observation that (some) linguistic categories follow specifiable distributions. Against this background, it is interesting that we still find differences in these plots, both by individual distributions and by topic and corpus. However, these are nowhere nearly as pronounced as those in the subclasses of verbs and nouns. What could explain the even larger variability in the realization of morphological, morphosemantic, and syntactic categories in our corpora then?

4.4. Priming and Self-Priming

Results so far have shown that cumulative corpus counts do not do justice to the internal distribution of the corpus, but that within a corpus, inter-individual variability needs to be accounted for. We further suspect that even a cumulative count of categories across an individual text marks a dimensionality

[19]More individual distribution plots are available through a Zenodo repository, 10.5281/zenodo.4752308. Only three out of four topics are available with corrected dependency labels in the current version of Falko, hence we are unable to show distributions for the *university* topic at present.

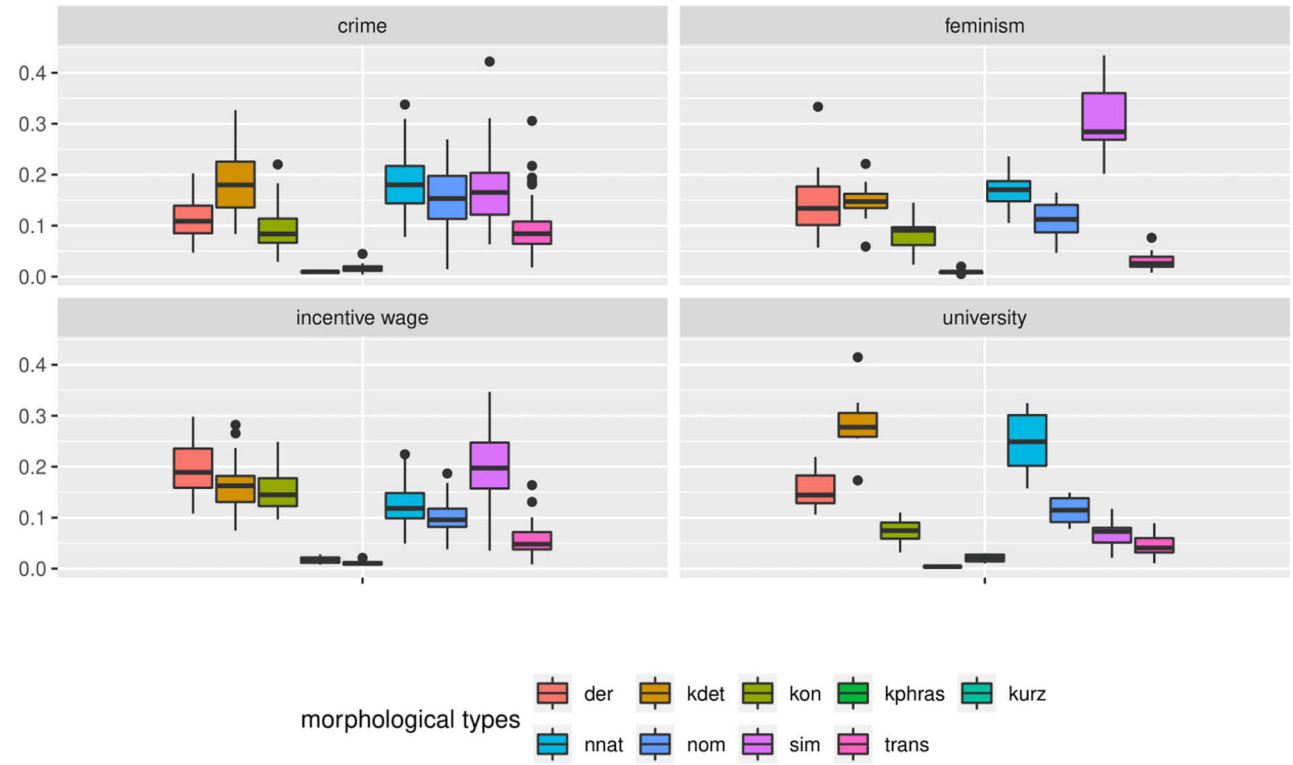

FIGURE 6 | The frequency of morphological types of nouns in Falko varies greatly when the distribution across individual documents is considered. Even some of the most frequent types, like simplex nouns *sim* in the incentive wage topic, barely occur in some texts. It also appears as if certain topics elicit more clearly defined distributions (*university* and *feminism*), while the overlap between boxes is much larger in the other two topics, particularly so in the *crime* topic. This may be due to higher random variability or an artifact from a lumping of different speaker, register, or text type typologies.

reduction that could hide some of the underlying dynamicity. We will therefore look into the role of priming in our phenomena. For this, we are going to take a closer look at distributions of specific morphological categories in course of the texts. In **Figures 11**, **12** we present data from a sliding window analysis of selected texts in Kobalt and Falko[20]. Each data point represents the number of elements of the respective category within a window of 50 tokens, for example 3 particle verbs within 50 tokens (words and punctuation). The first window spans tokens 1-50, the second 2-51, the third 3-53, and so on. There are *text length - 49* windows for each text.

If a category occurs once, the count stays at one until the windows have slid by its first occurrence. Thus, if a category occurs several times, the peak remains until the window slides past the *first* occurrence. A peak can persist over many windows if the first occurrence drops out but is replaced by another occurrence at the higher token end of the window. If a category is distributed equally or irregularly over the text, the line should be erratic: it is counted once or twice, then drops out, then occurs once or twice again. In many cases, however, we find peaks of five or even six occurrences of a category within a 50 token window[21]. This can be due to lexical repetition/recurrence, which is why we provide the number of unique lexemes for each category within the windows (marked dark blue in the plots). For most data points, their total occurrence overlaps with the number of lexemes, i.e., each occurrence represents a separate new word, not the repetition of previous words within that window[22].

Lexical recurrences would indicate lexical priming. Overall, we do not find strong evidence for this, although there are some cases. If there are many lexically diverse occurrences, that can indicate structural priming: once participants start using a structure, they stick with it, until they prime themselves to another category. We see strong evidence for this in the case of morphosemantic verb categories in Kobalt in **Figure 11**. Each row represents the four morphosemantic categories (particle, prefix, simplex, and support verbs) of an individual speaker,

[20]There are too many plots to present legibly in this paper. All remaining plots can be found along with the scripts for analysis in a Zenodo repository under 10.5281/zenodo.4752308.

[21]The number of tokens for each window is chosen arbitrarily. Since the windows overlap, no information is technically lost in smaller or larger windows. However, if the windows become too large, two peaks can be bridged, suggesting ongoing activation where in fact, there is a slump. If the windows are too small, peaks never reach levels higher than two or three, potentially clouding existing activation. A more exact calibration of this measure remains for future research and should be conducted in alignment with psycholinguistic research.

[22]We did not account for repetition across the whole text, but plan to do so in future research.

FIGURE 7 | The variety of distributions of noun types is even more striking when displayed by individual document. Even documents pertaining to the same topic vary greatly.

with data points representing the number of occurrences of each category in each window, similarly to a time series plot. Several texts show high peaks of a category, for example up to five prefix verbs within a 50-token window in texts DEU_002, DEU_017, and DEU_009. Perhaps even more intriguingly, it appears that there is a progression between forms, i.e., that speakers peak in one category and then move on to the next. This happens for instance in DEU_002, which features a range of windows with 3-4 particle verbs, followed by two peaks in prefix verbs; or in DEU_017, which begins with a number of prefix verbs, then introduces three particle verbs within a small number of windows—which are also the only three particle verbs in this text—and then returns to a peak in prefix verbs. Simplex verbs show more erratic curves, which might be due to their overall higher frequency or due to category conflation (perhaps certain types of simplex verbs prime for similar types that cannot be distinguished under the general *simplex* label). However, even simplex verbs interact with the other curves, for example in DEU_013, where the text begins with a high number of simplex verbs, which then make room for a peak in particle verbs, and then returns to a second peak in simplex verbs.

Figure 11 also shows that not all speakers are equally susceptible to clustering effects in morphological structure: DEU_007 does not show striking effects in particle, prefix, or support verbs; and DEU_018 shows nearly parallel curves for particle and prefix verbs, peaking twice at 3 vs. 4 occurrences respectively within a small range of windows. The number of unique lexemes closely follows the curves in all categories except simplex verbs in nearly all cases (with the exception of some particle verbs in DEU_002 and DEU_009). This suggests that the differences in the proportions of subclasses of verbs do not stem from different degrees of lexical richness of repetitive style. However, this does not conclusively mean that all forms are primed morphologically (structurally). It is possible that there is partial lexical priming through either the verb base or the prefix or particle, i.e., paradigmatic lexico-structural priming. This lies outside of the scope of this paper and will be treated separately in future research.

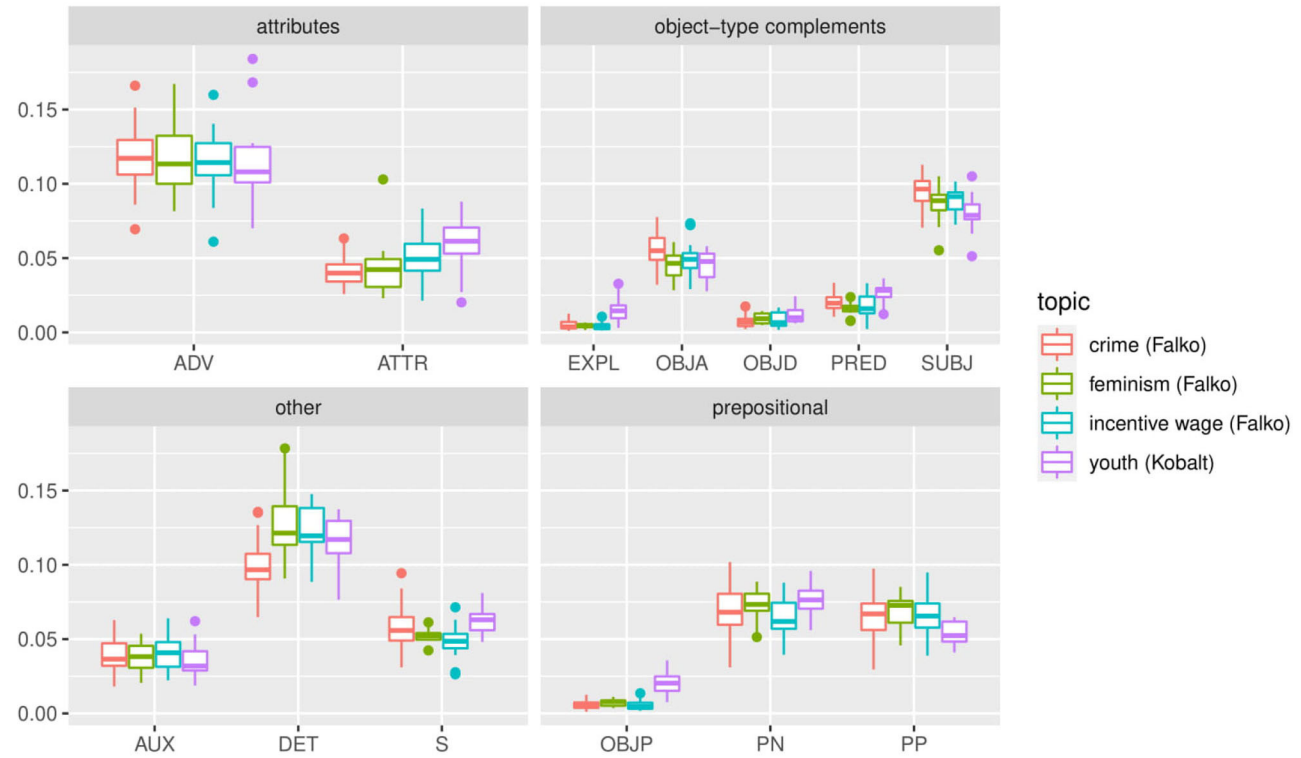

FIGURE 8 | Distribution of frequent dependency types in Falko and Kobalt relative to text length. Dependencies included: adverb and attribute (attributes), prepositions in free PPs and prepositional objects, nominal compliments to PPs (prepositional), subjects, predicates, accusative and dative objects (object-type complements), determiners, lexical, and auxiliary verbs (other). Although frequencies are overall more stable compared to morphological subclasses of verbs and nouns, there still are noticeable differences by topic/corpus (more attributes, prepositional objects, lexical verbs (S) in Kobalt; fewer determiners, more subjects in the *crime* topic in Falko. This might be due to differences in lexical choice or in register realization.

While there appears to be convincing evidence for priming or clustering effects for verbs in Kobalt, the case is more complicated for nouns. First of all, nouns vary more than verbs, both in the lexicon and each text, so that it is more difficult to set a baseline for when to assume a priming effect—each noun will belong to a morphological category, and since there are many, a number of each is to be expected in each window. Secondly, noun morphology is less transparent than verb morphology, and for some categories, structural properties are very abstract. This is the case for example in compounding, where the structure consists of only the combination of two words and headedness; or in transposition, where the structure is the use as another syntactic category rather than changes to the word itself. Unlike this, complex verb morphology, at least in the case of particle and prefix verbs, has a more distinct and obvious shape that speakers are likely more aware of (prefix/particle + base; plus phonetic features) or from which it is easier to draw connections to other forms. Compounding seems less restricted, it is hard to tell whether the form [noun + noun] was primed from a single noun or a compound. This requires a more detailed and qualitative analysis, which we will provide in a separate paper at a later time.

A clustering of categories can also be due to the coordination (listing) of elements, which is typical of some topics in Falko. For example, in the *university* topic, participants frequently mention a number of university programs such as biology, chemistry, psychology, etc., which in German tend to be of neo-classical origin (labeled as non-native). It is difficult to distinguish between this case and structural priming in less obviously related contexts without taking more qualitative evidence into account; and even where the evidence suggests one thing, there is no way to exclude structural priming effects in those lists—after all, it is possible that the list was provided, or at least extended, due to chained activation of similar lexemes.

In spite of these limitations, we suggest that there are potential cases of both self-priming and other-priming by the prompt in Falko noun morphology. We chose transpositions as our example here for self-priming. In **Figure 12**, the author of text dew07_2007_09 produces a series of transpositions with a peak at the beginning and several recurrences of this morphological noun type throughout the whole text. This distribution fits well with the observation that priming effects decrease with increasing distance from the prime. A similar distribution can be seen in dcs007_2007_10, whereas dhw022_2007_06 and dhw015_2007_06 exhibit constant recurrences of transpositions.

The usage of non-native nouns in the university topic subcorpus of Falko is an example for other-priming. The prompt

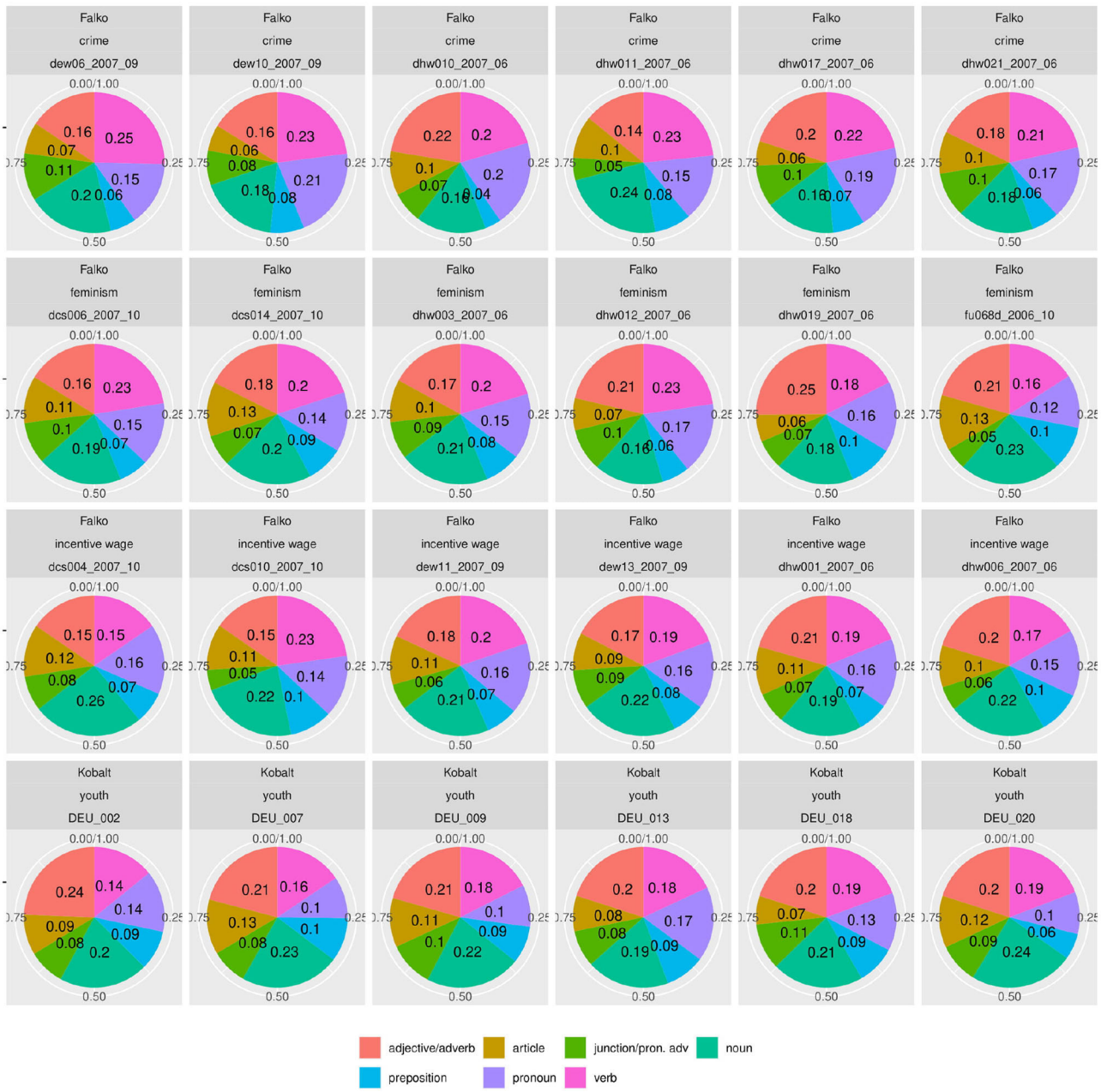

FIGURE 9 | Individual part-of-speech distributions in Falko and Kobalt are much more homogeneous than the morphological, syntactic, and morphosemantic subclasses of verbs and nouns.

for these texts contains complex nouns of the neo-classical word formation type (labeled as non-native) and primes the usage of other nouns of the same type. This can be seen in the numerous peaks for non-native nouns in the same plot, texts fu081d_2007_10 and fu082d_2007_10. Crucially, the dispersion of peaks indicates that the effect is not due to mere listing of non-native words within a single window.

In the case of dcs007_2007_10, we also find a similar pattern to the Kobalt data, namely the clustering of a category type in one part of the text vs. another in another part, with transpositions peaking earlier in the text than non-native nouns.

Our results show clearly that a cumulative account even of individual texts still masks intra-individual, or procedural, variation that occurs in peaks that in several cases shift or alternate between categories. While it is in principle possible to analyze our syntactic categories in the same way, there are some stricter limitations to both the necessity and the clarity of the analysis. Since syntactic elements appear to converge to

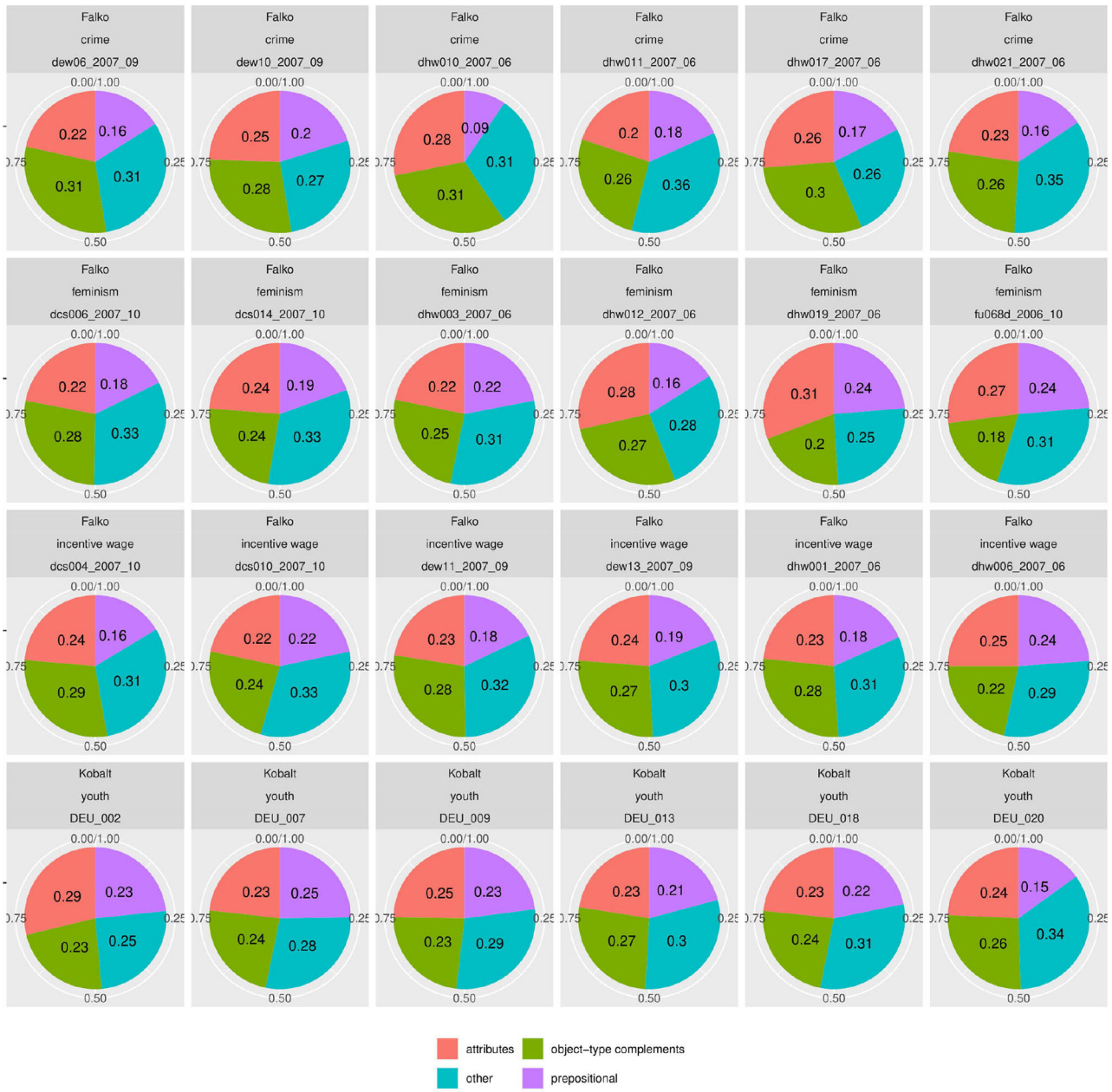

FIGURE 10 | Individual distributions of frequent dependencies in Falko and Kobalt documents. Dependencies included: adverb and attribute (attributes), prepositions in free PPs and prepositional objects, nominal complements to PPs (prepositional), subjects, predicates, accusative and dative objects (object-type complements), determiners, lexical, and auxiliary verbs (other). Proportions of dependency types to one another are much more homogeneous than morphological subclasses of elements. One exception is dhw_010_2007_06 (3rd from the left in top row), which contains only 9% of prepositional dependencies (prepositions and nominal complements to PPs).

a higher degree between speakers, cumulative counts of those are less problematic at least methodologically—if speakers can be expected to level out across text even in texts of divergent length, this would imply they also level out in shorter spans, and hence cumulative counts are less misleading overall. At the same time, accounting for syntactic priming by category is theoretically more complicated. This is due to the same reason as stated above for noun morphology, namely their high similarity (every noun is a noun, and they tend to occur frequently—what could provide certainty that this is due to priming?) and coordination (some participants like to list activities of a similar kind, like "reading books, magazines, or the newspaper," in which three accusative objects would occur within a very small window. Whether this should be considered

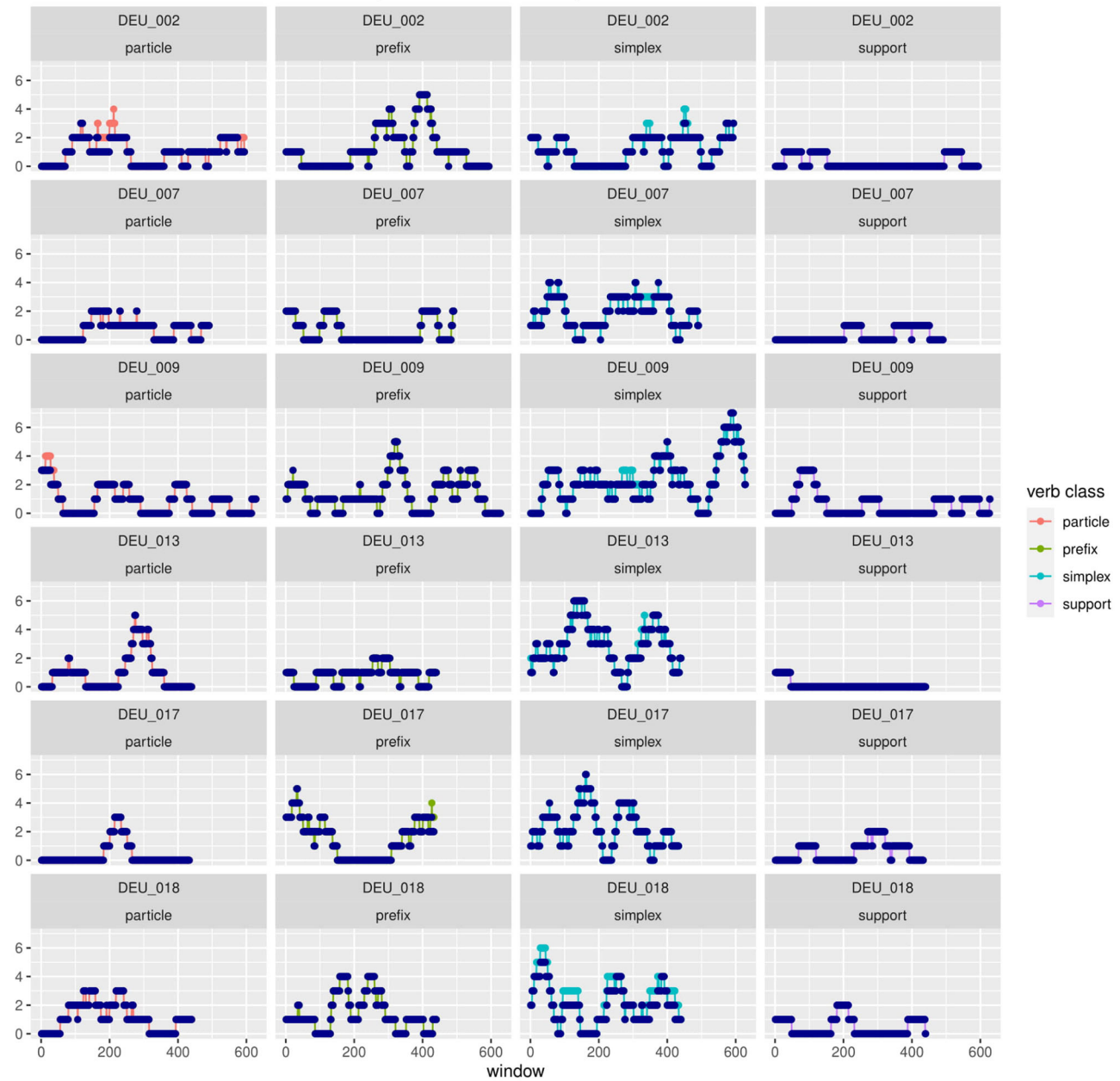

FIGURE 11 | Strong clustering effects can be seen for prefix (DEU_002, DEU_009, DEU_017) and particle verbs (DEU_013, DEU_017). However, not all speakers exhibit those to the same degree (DEU_007). Lexeme repetition or α-priming does not appear to play a significant role.

priming is unclear.) The charm of morphological priming is that words can be diverse within the same morphological category. It appears less likely that a participant will intentionally reuse the same category or coordinate several words of the same morphological category, but not the same lexeme, in the same way as a syntactic construction would allow. We will not exclude that possibility, but we will leave it for future research.

5. DISCUSSION

We began with the observation that L1-speaker data, aside from stratified or situational variation, is often conceptualized as a more homogeneous baseline in learner corpus studies, against which learner language is modeled as more diverse. While there has been a general paradigm shift in multilingualism research that models native speakers as less homogeneous than it used to,

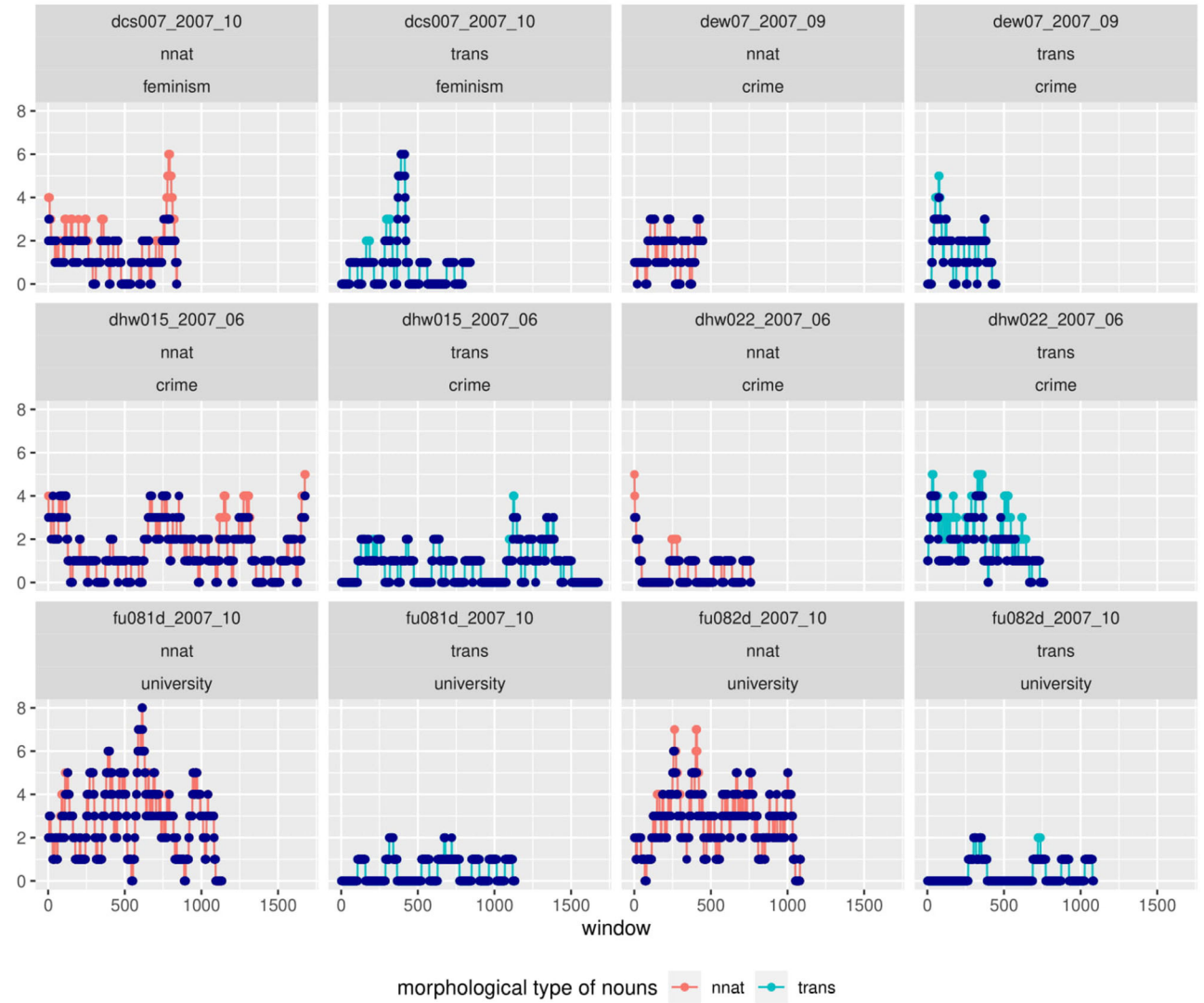

FIGURE 12 | A sliding-window analysis of nouns in Falko shows the clustering of categories in some, but not all texts. Morphological categories of nouns can strongly interact with topic, as in non-native nouns in the *university* topic. The repetition of lexemes within windows (suggestive of α-priming) is less likely than the introduction of new lexemes of the same type.

this paradigm shift is based on a prism refracting the formerly monolithic model of native language into a large number of diverse group memberships, not unlike the intersectional approach to society in general. For example, native speakers are not a homogeneous group if attributed as such by country of residence or exposure to the target language alone. They may differ by a number of language-external factors (such as age, region, or socio-economic status) and several language-peripheral factors (such as reading experience and linguistic aptitude), some of which may be explanatory in the diversity of use. There are also clearly influences of linguistic environments that trigger one linguistic realization over another, as if setting switches probabilistically and independent or only partially dependent on other characteristics of the speaker.

However, this is *still* a stratified view. We maintain that even approaches accounting for such systematic differences do not do justice to the full variability present in native speaker data. Quantitative and qualitative differences are strongly expressed even in the analysis of a highly homogeneous group of speakers, but this appears to be the case for some linguistic levels more than others. In other words, the speakers in our corpora were selected to be as homogeneous as possible, limiting participation to the literal same classroom in the case of Kobalt, yet *still* we find quantitative differences in morphology, but relative homogeneity in syntax. Both the high degree of variance in German morphology and the divergence between degrees of variance between linguistic levels is to our best knowledge previously undescribed.

We have further shown that all except the vanishingly rare categories are equally subject to high degrees of variability, even those that would be considered a prototype or baseline category such as simplex verbs or nouns; and that even relatively coarse aspects of the total distributions, such as the order of categories by frequency or even the category ranking highest by frequency, could not be determined across speakers in our corpora. This is in spite of highly controlled elicitation conditions as well as identical prompts between participants.

This is relevant in the context of a growing interest in morphological complexity in SLA. It is also highly relevant in the context of learner corpus and other usage-based studies, that largely work from a contrastive paradigm even if they do not explicitly state this, but as is evident from the methodology they apply.

Studies that concentrate more on the nature of presumed input of learners in most cases also do this in contrast to a native control group of some sort. Linford et al. (2016), where the make-up of the control group is one of the independent variables, do not only consider a global and local corpus as comparison, but entertain the possibility of a control group of other learners[23]. The crucial point, however, is that for none of the group-data—be it from the local or global corpora—inter-individual variability is reported. Geeslin et al. (2013) and some of the references therein are an exception insofar as they do report standard deviations of their control group. However, our data further shows that native speakers do not simply differ in their realization or non-realization of a binary category, but *in the whole composition* of their morphological subclasses for nouns and verbs, but *the same speakers* do *not* differ to a comparable degree in higher-order and more systemic syntactic compositions.

We have further shown that even the degree of intra-individual variation can be high and appears to follow systematic patterns organized by procedural effects. Variable degrees of intra-individual variation would be expected to transcend into variable degrees of inter-individual variation. If, for example, some participants prime themselves to the use of particle verbs, they will use them more overall than those who are less susceptible to self-priming or who do not happen to use a particle verb before they finish their text. This highlights the non-ergodicity, or path-dependence, of the writing process. However, in corpus linguistics, corpora are largely treated as static, non-dynamic data, with perhaps the exception of dialogue corpora and the smaller number of corpus-based priming studies that are available to date. Our data suggests that these aspects may deserve more attention in the future.

One of the reasons for why inter-individual differences of this scale even among (theoretically) homogeneous control groups have not attracted more attention so far may be in the syntactic and/or lexical focus of much of corpus research: Gurzynski-Weiss et al. (2018) have shown that the preference for a specific form of subject expression in L1-Spanish correlates with grammatical context and situational setting, while Linford et al. (2016) report distributional differences in context-integration between their learners and control groups. Since these contextual variables (level of attainment being one of them) lend themselves quite well to explaining the observed differences, there seems to be no need to delve deeper into inter-individual differences on the side of the control group. The morphological phenomenona observed in our data, on the other hand, evade the same kind of explanation. All of our speakers realize nearly all of the forms, and where they do not, it is clearly not a function of attainment.

In conclusion, various usage-based models think of L1 frequencies as representations of relevant quantitative properties of the target language, and frequently interpret L2 frequencies as over- or underuse. If subclasses are not equally distributed across native speakers, like our data shows for verb and noun morphology, this perspective needs to be expanded to include inter-individual and perhaps even intra-individual differences in L1. We will briefly discuss methodological implications for learner corpus studies and theoretical issues that arise for cognitive/usage-based models of SLA.

5.1. Methodological Implications

We began this paper by stating that in many learner corpus studies, native speaker data is used as a control group for comparison with learners. In the contrastive paradigm, higher or lower frequency of occurrence of various linguistic elements in learner data is frequently viewed as evidence for a learner's target language competence. This methodologically implies native speakers as somewhat idealized carriers of the target language that converge both qualitatively and quantitatively, even where the research paradigm theoretically states otherwise. Learners are naturally presumed to exhibit higher variability, as are bilinguals in general (Seton and Schmid, 2016, 341).

This does not match with our data of a (theoretically) very homogeneous group of native speakers of German. We conclude that it is therefore important to refrain from comparing groups by cross-corpus means without further investigation of variance and distribution, and we should not presume native speaker homogeneity *across linguistic categories*. For a valid group comparison, the distribution within the group must be both (a) known and (b) comparable. We cannot rely on median or mean values as long as the variability tendency of the phenomenon at hand is unclear, which means that for any corpus statistic, the inter-individual comparison *must* be accounted for and reported. This can complicate matters, especially where individual contributions are not trivially attributable or where the research questions requires the consideration of rare phenomena that do not always manifest in the writing of every individual. However, a quantitative analysis is only meaningful if we understand the underlying, expected, and measured distribution adequately.

This is especially relevant where statistical models are employed, because those typically rely on certain assumptions that may not be met by vastly variable within-group distributions.

[23] While this is quite plausible, the way it is operationalized is problematic: the learner data taken as verum group is simply added to the native bilinguals' group data that already serves as one of the control groups, and this "supergroup" is then entered as the third control group, albeit neither independent from the learners nor the native bilinguals.

The most basic assumption of statistical models is that phenomena have a probability, which in frequentist statistics is defined as the outcome of each factor in terms of relative frequency of an infinite series of random experiments. In other words, if I draw samples from the same population a large number of times, over time, the relative frequency for each state (each morphological subclass, for example) should stabilize, i.e., converge to an idealized value, which is the probability. If it does not, this can be due to the phenomenon not having a stable probability: it may be too dynamic, e.g., driven by intention, the invisible hand of cognitive and procedural factors such as priming, or a combination of those two with more general frequency patterns. In that case it might best be understood as a complex dynamic subsystem (individual grammar/*parole*) within a larger complex dynamic system (speaker group language/*langue*)[24]. If a phenomenon does not have a stable probability, statistically inferring from a sample to a population is meaningless (see Shadrova, ress, for a more in-depth argument).

In our data, speakers do not converge to one another in their use of more fine-grained categories in a single text, while they do appear to converge (within a range) in some other categories. Would more data resolve the issue? Do speakers converge to one another, i.e., follow general frequency patterns in the use of subclasses of verbs and nouns, but a single text does not provide a sufficiently large speaker-specific sample? Do they follow different, but contextually stable frequency distributions, for example by text type or register, and would these converge between speakers? Do they not converge to one another, but stabilize in their own frequency patterns—i.e., are there idiosyncratic frequency distributions for each speaker? Or is there simply no convergence between or within speakers, i.e., should we allow for random fluctuation within a range of between 15 and 66% of simplex verbs in Kobalt (DEU_001 vs. DEU_011) and 0 and 53% of modal verbs (DEU_018 vs. DEU_005)? In this case, we would have to accept that a simplistic groupwise comparison of the phenomenon based on frequencies is pointless. Statistics is a scientific belief management system designed to filter the signal from a noisy (variable) environment. However, measurements that hit both floor (0 occurrences) and what could be considered ceiling (53% modal verbs) complicate the analysis. It is possible that looking more into the shapes of the distributions and their interactions with other category distributions would yield clearer results. Either way, if native speakers, i.e., target language carriers, use between zero and as many modal verbs as reasonably possible, the precise mapping and comparison of learner data to this raises methodological questions.

More importantly, our results raise linguistic questions: what is going on in the language of speakers that do not use any modal verbs? How do they construct modality instead? What is different in the language of speakers who barely use auxiliaries, copula, or constructional verbs, but many modal verbs? How does their language differ from all the other speakers in the corpus? Should we attempt to capture morphosyntactic speaker profiles instead of individual varieties? These questions in turn trigger methodological considerations that go beyond the question of adequate statistical description and analysis.

5.2. The Role of Frequency in Usage-Based Accounts

As has been briefly discussed in section 2, usage-based accounts of language acquisition and production make a strong point of emphasizing the role of frequency in the input. This applies to the whole range of the continuum from syntactic constructions to individual words and word co-occurrences (Bybee and Hopper, 2001; Gries and Wulff, 2005; Ellis and Frey, 2009; Ellis, 2012; Goldberg, 2013; Diessel and Hilpert, 2016; Hilpert, 2017; Gries, 2019, and many others). The idea is that speakers are sensitive to frequency distributions because frequencies of linguistic elements acquire neuronal correlates by means of entrenchment (strengthening of neuronal pathways through repetition, resulting in effortless reproduction of the entrenched frequencies). An element that is frequently heard or seen will be frequently produced and more easily recognized. They also make the case that all linguistic units exist on a continuum of form-meaning pairs that in principle are learned in the same way, or that "it's constructions all the way everywhere" (Boogaart et al., 2014, 1).

Our data provides challenges to this account. It has been collected from participants from homogeneous backgrounds – to the extent that our high school students would be faced with similar books at school, share significant amounts of daily conversation and a similar social environment in many ways. Still, they either do not arrive at the same distributional abstractions, or do not reproduce those abstractions in the same way. This means that either (a) frequency in entrenchment is not automatically mirrored in production, (b) that there is another factor determining frequency distributions that is currently being overlooked (such as latent register differences between texts) or simply (c) that not all constructions are entrenched with frequency. But what would that imply for "constructions all the way down" (Goldberg, 2006)?

In our data, we do not find the same divergence between individuals for some higher-level syntactic relations, parts of speech and dependencies. It is possible that this is not an effect of abstraction/concreteness, but one of relational function: unlike verbs or nouns of different morphological types, the different dependency types or parts of speech form a system. Thus the total 100% of all dependencies in a text are mutually interdependent to a large degree—one can often not easily add a verb without also adding nouns, or a noun without also adding a determiner/quantifier etc[25].—while the elements tallied in the other categories are mutually independent (using an extra prefix verb does not grammatically enforce the next particle verb, for

[24]For *langue* and *parole*, see Saussure ([1916]1983). For the modeling of language as a complex dynamic system (see Ellis, 2006; Five Graces Group et al., 2009; Lowie and Verspoor, 2019), among others.

[25]Except in the coordination of lists of activities, as in *they drank tea and danced and laughed*. However, even this will sooner or later trigger nouns: *. . . and played guitar and told stories. . .*, and also cannot be continued ad infinitum in a realistic context due to limitations in processing as well as its communicative pointlessness.

example). A system is defined by the mutual interrelationships of its elements (Mesarovic, 1964), producing a latent structure which might be accountable for stable frequencies. It is possible that speakers are not as much sensitive to *frequencies* as they are to *proportions* within a (sub-)system, or in other words that frequency is an epiphenomenon of structured inventories of signs, not a feature of the signs themselves[26]. This would go against the idea of equality of all linguistic signs and categorizations as it is prominent in usage-based accounts ("constructions all the way down" Goldberg, 2006, 18). For a valid quantitative statement, one would then need to define the respective subsystem first.

One relevant question in this regard is whether the differences in morphological category distributions could be explained by looking at lexical, rather than morphological, frequencies. Theoretically speaking, morphologically complex words could in principle be realized without taking note of their complexity (as chunks or words without deeper analysis). While it is necessary to have an abstraction over forms for felicitous productivity, this is not necessary for the plain use of form. One could argue that it is possible that complex verb forms go largely unanalyzed in some or most speakers—that they are fully lexicalized and their distributions merely an epiphenomenon, that "meaning overrides frequency" (Jolsavi et al., 2013). However, as is frequently argued in usage-based approaches, schemas must be accessible in lexicalized forms, too, since productivity and generativity is considered to emerge from usage, and grammar from the use of lexemes (Booij, 2013; Zeldes, 2013; Hilpert, 2019, and others). If the schema is present in all use, and frequency is part of the schema, would we not expect less variable distributions between speakers?

With respect to the the data model and analysis, if word frequencies were stable, so would be morphological frequencies, because words do not change their mophological class. A higher level of abstraction would always reduce noise due to the loss of individuality of the lexemes. If anything, morphological categorization should level out the variance (higher dispersion) from more granular categories such as lexemes. There is also no evidence for lexical convergence or considerable overlap between authors in our corpora.

The problems with this perspective run deeper, though. Statistical approaches to word frequencies as quantifications of the lexicon in use have a long history in corpus linguistics (Baayen, 2002; Stefanowitsch and Gries, 2003; Gries and Wulff, 2005; Gries, 2013, 2019; Brezina et al., 2015, and many others). However, there are major mathematical and philosophical flaws. If word frequencies are not stable, i.e., stationary, and ergodic, i.e., path-independent (unaffected by factors such as priming or intention), they cannot be validly used for statistical computation. This is because all frequentist statistics relies on the central limit theorem, which does not hold true in systems that are non-ergodic or not stationary (Shadrova, ress; Schmid, 2010; Koplenig, 2017). There is mathematical research suggesting that language is overall non-ergodic (Dębowski, 2018). This could potentially be tackled by defining ergodic subsets. However, there is also evidence that even large corpora may not be stationary (Piantadosi, 2014; Shadrova, 2020) shows that for Kobalt, there is barely any lexical overlap between texts.

Most importantly, however, the way words are distributed in natural language makes word frequencies largely an artifact of corpus size. While there are groups of words that tend to occur more frequently, highly frequently, and so on, they escape any precise or meaningful quantitative categorization. Words as they occur in corpora follow a long-tailed distribution which is marked by a few highly frequent and some less frequent words, and a very large number of words that occur only once (*hapax legomena*). The larger the corpus, the more hapaxes. This is true of individual text and text corpora equally. For most words, their frequency thus is 1 divided by corpus size. There is no evidence that word frequencies are stable (stationary) in any corpus size. If it were, there could be no productivity, because all new words would take up space. It is clear that word frequencies can fluctuate more systematically (some disappear; some disappear, then reappear), however, such fluctuations are unpredictable beforehand. It is the statistical equivalent to rolling a die with a changing number of sides. The same is not true of morphological categories, which at least synchronically show some stability and a level of certainty of occurrence. While not every one of our participants uses all morphological categories, most classes are well represented and pooling only a few texts leads to good coverage of all classes. The same is far from true for lexemes in *any* corpus size.

It is of course possible that other factors can explain the divergence in individual distributions of classes of verbs and nouns in native speaker writing in these corpora. It might be a matter of aptitude or experience, style, or cognitive biases such as priming. Even then, usage-based linguistics needs to clarify the role of frequency and variance across linguistic categories in interaction with these factors. This is necessary for descriptive adequacy—if we observe heterogeneity in frequency realizations between native speakers, our theoretical models should capture this fact. It is equally necessary for explanatory adequacy—something makes speakers arrive at different frequency realizations in some, but not all categories, and usage-based theory at present does not provide a mechanism for this.

The divergence between category frequencies in production is also relevant for the question of input. Since the data we collected is semi-naturalistic—it has been collected for a linguistic purpose, but it is not unlike tasks that students are faced with in high school or college in Germany—we can assume that this is a realistic production scenario. If it is a realistic production scenario, it must also be a realistic input scenario: if speakers can choose to use simplex verbs between 15 and 66% in a text, then those who read those texts are equally confronted with such differences. While in a corpus, frequency may or may not level out, speakers outside of corpus linguistics are rarely confronted with a corpus to read. How are speakers not confused in their entrenchment of the frequency of morphosemantic constructions such as "particle verb" or "simplex verb" if those frequencies fluctuate by such vast amounts between texts? What does it say

[26] A similar suggestion with respect to coselectional constraints on verb-argument structures has been made by Shadrova (2020, 264-265).

about a phenomenon if it allows for high degrees of seemingly random fluctuation?

We will not exclude the possibility that there is some stratified variation between speakers in our corpus that we have not been able to account for yet. Wherever data occurs with high variation, the possibility of subgroups, such as a speaker typology by preference or style of expression, should be considered. This remains for future research and modeling. For this analysis, we chose to look into more procedural factors, which tend to be less in focus in corpus linguistic research. Our analysis is consistent with a priming-based explanation of at least some of the variability in our corpus. If the occurrence of one particle verb primes for three or four more such verbs, this would have great impact on the overall distribution in a text of 600 tokens, for example. It is plausible to assume that we find less variation in the more global syntactic phenomena due to varying degrees of susceptibility to priming. Global syntactic categories may be largely fixed through inherent constraints of the system, while morphological and other more fine-grained categories may be more susceptible to priming. Yet others may be subject to more free choice or control through speaker intention, resulting in stylistic choices. Such effects may differ by various factors, such as speaker aptitude, writing experience, or different register perception and knowledge.

Of course, this is a slippery slope. It might be tempting to suggest that fluctuations in frequency, whether they stem from preferences or priming, are a "performance" issue similar to how traditional generative grammar has declared ungrammatical sentences out of scope of syntactic research. This would miss out on a chance to learn about deeper structural differences between those categories that allow for fluctuations vs. those that do not appear to do so, which has multiple repercussions on procedural (connectionist) theories of language learning, production, and productivity. It would also pose challenges to the development of more adequate models for prediction and analysis of results in quantitative corpus studies. Most importantly, it would introduce a major inconsistency into constructionist models of language acquisition, because it would define frequency as both *relevant in acquisition and reception* and *irrelevant in production*, which is logically inconsistent, since reception depends on production.

We would like to emphasize that none of this is to say that there are no differences between L1 and L2 usage of morphological categories, or that "everything is just very, very diverse and cannot be captured". Rather, we argue for precise modeling from factors already available in many corpora, namely a document-wise analysis and consideration of a view of text as process. Native speaker writing is more complex than is frequently accounted for at present, and a more comprehensive view would emerge from an adequate representation of methodological decisions in theoretical modeling as well as vice versa.

5.3. Conclusion and Future Research

In this paper, we have presented data from two task-specific German L1 corpora that were initially collected as control corpora for second language acquisition studies. We have shown that in these two corpora, which are carefully compiled and controlled by a number of factors such as text type, writing conditions, participant background, and prompt, native speakers show high quantitative variance in the distribution of morphological subclasses of verbs and nouns, both between and within speakers. We have also shown that part-of-speech and syntactic dependency distributions do not appear to be subject to the same variability. As our morphological data suggests, it appears that even the gratifying departure from the assumption of native speaker homogeneity as it is represented in variationist and multilingualism-centered perspectives is not yet taking things far enough.

Future research needs to clarify the stability of the degrees of variance we find in native speaker writing for different levels of linguistic description. Do speakers, for example, show stable and persistent individual distributions of morphological types in verbs and nouns, or is high variance triggered through priming? How much of this is driven by intention/rhetorics, and how much is cognitively biased? What is the role of speech rhythm/accent patterns and phonetic priming, and what is the role of semantic priming in the repetition of (seemingly) abstract structures?

Our results highlight the importance of accounting for inter- and even intra-individual variance in corpus studies. In fact, some phenomena show such high degrees of variance that a quantitative comparison without further specification of the model appears pointless. This is crucial for quantitative studies—in order to study differences between language learners and native speakers, we need to know which phenomena allow for a meaningful quantitative comparison and which ones do not. Beyond this empirical implication, theoretical questions arise with respect to the role of frequency and item distributions that have traditionally been emphasized in usage-based linguistic theory, both in language learning and production in L1 and L2. If speakers produce vastly different quantitative outputs, then the role of quantitative entrenchment and its repercussions on language in use becomes much less clear and its centrality and implication as a lever in language learning may need to be reassessed at least for some linguistic levels.

Finally, if syntactic units are easy to quantify and converge quickly, while morphological units show different behaviors, and lexical material is even more difficult to grasp in mathematically valid ways, the idea of "constructions all the way down" (or "all the way everywhere") should be discussed in a more differentiated manner. While all these linguistic elements can be conceptualized as signs or form-meaning pairs on some level, the mechanisms facilitating their acquisition and production appear to differ at least with respect to their sensitivity to frequency and their (in)equation of frequency and entrenchment.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.zenodo.org/record/3584091; 10.5281/zenodo.4752308.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

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

FUNDING

Research Unit Emerging Grammars in Language Contact Situations: A Comparative Approach, FOR 2537, SH 1685/1-1, LU 856/16-1, 313607803, (AS, AL, and PL). SFB 1412 Register, funded by Deutsche Forschungsgemeinschaft, 416591334 (JL and AL). Crosslingual Language Varieties, funded by Deutsche Forschungsgemeinschaft, LU 856/13-1, 398186468 (SS and AL).

ACKNOWLEDGMENTS

A special thank you is dedicated to our colleagues Felix Golcher and Martin Klotz for their fruitful comments. We would also like to thank our student assistant Roodabeh Akbari for her active support. We would like to express our gratitude to three reviewers, whose constructive criticism and valuable advice has greatly helped to improve this paper. Any remaining inaccuracies are of course attributable to us. We acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.716485/full#supplementary-material

REFERENCES

Ädel, A. (2015). "Variability in learner corpora," in *The Cambridge Handbook of Learner Corpus Research*, eds S. Granger, G. Gilquin, and F. Meunier (Cambridge, UK: Cambridge University Press), 401–421.

Arroyo, J. L. B., and Schulte, K. (2017). Competing modal periphrases in spanish between the 16th and the 18th centuries: a diachronic variationist approach. *Diachronica* 34, 1–39. doi: 10.1075/dia.34.1.01bla

Artstein, R., and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Comput. Linguist.* 34, 555–596. doi: 10.1162/coli.07-034-R2

Baayen, R. H. (2002). *Word Frequency Distributions, Vol. 18*. Dordrech; Boston; London: Springer Science & Business Media.

Backus, A. (2021). "Usage-based approaches," in *The Routledge Handbook of Language Contact, Routledge Handbooks in Linguistics, Chapter 6*, eds E. Adamou and Y. Matras (London: Routledge), 110–126.

Bates, E., Dale, P. S., and Thal, D. (1995). "Individual differences and their implications for theories of language development," in *The Handbook of Child Language*, eds P. Fletcher and B. MacWhinney (Oxford: Blackwell), 96–151.

Bayley, R. (2019). "Variationist sociolinguistics," in *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press.

Berman, R. A., and Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: a developmental paradox. *Discourse Process.* 43, 79–120. doi: 10.1080/01638530709336894

Bernaisch, T., Gries, S. T., and Mukherjee, J. (2014). The dative alternation in south asian english (es): modelling predictors and predicting prototypes. *English World Wide* 35, 7–31. doi: 10.1075/eww.35.1.02ber

Bestgen, Y., and Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: an automated approach. *J. Second Lang. Writing* 26, 28–41. doi: 10.1016/j.jslw.2014.09.004

Biber, D. (1993). Representativeness in corpus design. *Literary linguist. Comput.* 8, 243–257. doi: 10.1093/llc/8.4.243

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguist. Linguist. Theory* 8, 9–37. doi: 10.1515/cllt-2012-0002

Biber, D., Egbert, J., Gray, B., Oppliger, R., Szmrecsanyi, B., Kyto, M., et al. (2016). "Variationist versus text-linguistic approaches to grammatical change in English: nominal modifiers of head nouns," in *Cambridge Handbooks in Language and Linguistics* (Cambridge, UK), 351–375.

Biber, D., and Jones, J. K. (2009). "Quantitative methods in corpus linguistics," in *Corpus Linguistics. An International Handbook, Vol. 2* (Berlin: De Gruyter Mouton), 1286–1304.

Birdsong, D. (2021). Analyzing variability in l2 ultimate attainment. *Lang. Interact. Acquis.* 12, 133–156. doi: 10.1075/lia.21001.bir

Birdsong, D., and Gertken, L. M. (2013). In faint praise of folly: a critical review of native/non-native speaker comparisons, with examples from native and bilingual processing of French complex syntax. *Lang. Interact. Acquis.* 4, 107–133. doi: 10.1075/lia.4.2.01bir

Boas, H. C. (2013). "Cognitive construction grammar," in *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.

Bonfiglio, T. P. (2010). *Mother Tongues and Nations*. Berlin: De Gruyter Mouton.

Boogaart, R., Colleman, T., and Rutten, G. (2014). "1. constructions all the way everywhere: Four new directions in constructionist research," in *Extending the Scope of Construction Grammar* (Berlin: De Gruyter Mouton), 1–14.

Booij, G. E. (2013). "Morphology in construction grammar," in *The Oxford Handbook of Construction Grammar*, eds T. Hoffmann and G. Trousdale (Oxford: Oxford University Press), 255–273.

Borer, H. (1996). Access to universal grammar: the real issues. *Behav. Brain. Sci.* 19, 718–720. doi: 10.1017/S0140525X00043582

Brezina, V., McEnery, T., and Wattam, S. (2015). Collocations in context: a new perspective on collocation networks. *Int. J. Corpus Linguist.* 20, 139–173. doi: 10.1075/ijcl.20.2.01bre

Brezina, V., and Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Lang. Res.* 35, 99–119. doi: 10.1177/0267658316643125

Bybee, J. (2002). Sequentiality as the basis of constituent structure. *Typol. Stud. Lang.* 53, 109–134. doi: 10.1075/tsl.53.07byb

Bybee, J., and Torres Cacoullos, R. (2009). The role of prefabs in grammaticization: How the particular and the general interact. *Formulaic Lang.* 1, 187–217. doi: 10.1075/tsl.82.09the

Bybee, J. L. (2013). "Usage-based theory and exemplar representations of constructions," in *The Oxford Handbook of Construction Grammar* (Oxford: Oxford University Press), 49–68.

Bybee, J. L., and Hopper, P. J. (eds.). (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins Publishing.

Cacoullos, R. T., and Travis, C. E. (2019). Variationist typology: Shared probabilistic constraints across (non-) null subject languages. *Linguistics* 57, 653–692. doi: 10.1515/ling-2019-0011

Chapman, M. D. (2016). *The Effect of the Prompt on Writing Product and Process: A Mixed-Methods Approach* (Ph.D. thesis). University of Bedfordshire.

Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. *Interlang. Stud. Bull.* 7, 103–117. doi: 10.1177/026765839100700203

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach*. London: Longman.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguist. Approaches Bilingualism* 2, 219–253. doi: 10.1075/lab.2.3.01dab

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* 178, 222–235. doi: 10.1016/j.cognition.2018.05.018

Davies, A. (2011). Does language testing need the native speaker? *Lang Assess Q.* 8, 291–308. doi: 10.1080/15434303.2011.570827

Davies, M. (2002). Un corpus anotado de 100.000.000 palabras del español histórico y moderno. *Procesamiento del lenguaje natural.* 29, 21–27.

De Clercq, B., and Housen, A. (2019). The development of morphological complexity: a cross-linguistic study of L2 French and English. *Second Lang. Res.* 35, 71–97. doi: 10.1177/0267658316674506

Dębowski, Ł. (2018). Is natural language a perigraphic process? The theorem about facts and words revisited. *Entropy* 20, 85–111. doi: 10.3390/e20020085

DeKeyser, R. (2012). Individual differences in native language attainment and their implications for research on second language acquisition. *Linguist. Approaches Bilingualism* 2, 260–263. doi: 10.1075/lab.2.3.03dek

Deshors, S. C., and Gries, S. T. (2016). Profiling verb complementation constructions across new englishes. *Int. J. Corpus Linguist.* 21, 192–218. doi: 10.1075/ijcl.21.2.03des

Diessel, H., and Hilpert, M. (2016). "Frequency effects in grammar," in *Oxford Research Encyclopedia of Linguistics*, eds M. Aronoff (Oxford: Oxford University Press).

Divjak, D., and Caldwell-Harris, C. (2015). "Frequency and entrenchment," in *Handbook of Cognitive Linguistics, HSK 39*, eds D. Divjak and E. Dąbrowska (Berlin: De Gruyter Mouton), 53–75.

Doerr, N. M. (2009). "Investigating "native speaker effects": toward a new model of analyzing "native speaker" ideologies," in *The Native Speaker Concept. Ethnographic Investigations of Native Speaker Effects*, ed N. M. Doerr (Berlin: De Gruyter Mouton), 11–46.

Dörnyei, Z. (2005). *The Psychology of the Language Learner. Individual Differences in Second Language Acquisition*. New York, NY: Routledge.

Dubois, S., and Sankoff, D. (2001). "The variationist approach toward discourse structural effects and socio-interactional dynamics," in *The Handbook of Discourse Analysis* (Malden, MA: Blackwell), 282–303.

Eckert, P. (2016). "Third wave variationism, in *The Oxford Handbook of Sociolinguistics* (Oxford: Oxford University Press).

Eckes, T. (2010). "Der Online Einstufungstest deutsch als fremdsprache (OnDaF): theoretische grundlagen, konstruktion und validierung," in *Der C-Test: Beitrüge aus der aktuellen Forschung, The C-Test: Contributions from Current Research*, ed R. Grotjahn (Frankfurt a. M: Peter Lang), 125–192.

Ehret, K., and Szmrecsanyi, B. (2016). Compressing learner language: An information-theoretic measure of complexity in sla production data. *Second Lang. Res.* 35, 23–45. doi: 10.1177/0267658316669559

Ellis, N. (2012). "Frequency-based accounts of second language acquisition," in *The Routledge Handbook of Second Language Acquisition*, eds M. Susan and A. M. Gass (New York, NY: Routledge), 193–210.

Ellis, N., and Frey, E. (2009). "The psycholinguistic reality of collocation and semantic prosody (2)," in *Formulaic Language: Acquisition, Loss, Psychological Reality, and Functional Explanations, Vol. 2*, eds R. Corrigan, E. A. Moravcsik, H. Ouali and K. M. Wheatley (Amsterdam: John Benjamins), 473–497.

Ellis, N. C. (1996). Sequencing in SLA: phonological memory, chunking, and points of order. *Stud. Second Lang. Acquis.* 18, 91–126. doi: 10.1017/S0272263100014698

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acquis.* 24, 143–188. doi: 10.1017/S0272263102002024

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Appl. Linguis.* 27, 1–24. doi: 10.1093/applin/ami038

Ellis, N. C., and Wulff, S. (2015). "Usage-based approaches to SLA," in *Theories in Second Language Acquisition-An Introduction, Second Language Acquisition Research Series, Chapter 5, 2nd Edn*, eds B. Van Patten, and J. Williams (London: Routledge), 75–93.

Eskildsen, S. W., and Cadierno, T. (2015). "Advancing usage-based approaches to L2 studies," in *Usage-Based Perspectives on Second Language Learning*, eds T. Cadierno and S. W. Eskildsen (Berlin: De Gruyter Mouton), 1–16.

Foth, K. A. (2006). *Eine umfassende Constraint-Dependenz-Grammatik des Deutschen*. Hamburg: Fachbereich Informatik; Universität Hamburg.

Gablasova, D., Brezina, V., and McEnery, T. (2017). Exploring learner language through corpora: comparing and interpreting corpus frequency information. *Lang. Learn.* 67, 130–154. doi: 10.1111/lang.12226

Geeslin, K., Linford, B., Fafulas, S., Long, A., and Díaz-Campos, M. (2013). "The L2 development of subject form variation in Spanish: the individual vs. the group," in *Selected proceedings of the 16th Hispanic Linguistics Symposium*, eds J. Cabrelli Amaro, G. Lord, A. de Prada Pérez and J. E. Aaron (Somerville, MA: Cascadilla Proceedings Project), 156–174.

Geeslin, K. L., and Long, A. Y. (2014). *Sociolinguistics and Second Language Acquisition-Learning to Use Language in Context*. London: Routledge.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, A. E. (2013). "Constructionist approaches," in *The Oxford Handbook of Construction Grammar*, eds T. Hoffmann and G. Trousdale (Oxford: Oxford University Press), 15–31.

Goldberg, A. E., Casenhiser, D. M., and Sethuraman, N. (2004). Learning argument structure generalizations. *Cogn. Linguist.* 15, 289–316. doi: 10.1515/cogl.2004.011

Graces Group, Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., et al. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59, 1–26. doi: 10.1111/j.1467-9922.2009.00533.x

Granger, S. (2002). "A bird's-eye view of learner corpus research," in *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching, Vol. 6 of Language Learning Language Teaching*, eds S. Granger, J. Hung and S. Petch-Tyson (Amsterdam: John Benjamins), 3–33.

Granger, S. (2005). "Pushing back the limits of phraseology: how far can we go," in *Phraseology 2005: The Many Faces of Phraseology*, eds C. Cosme, C. Gouverneur, F. Meunier and M. Paquot (Louvain-la-Neuve: CECL), 165–168.

Granger, S. (2015). Contrastive interlanguage analysis: a reappraisal. *Int. J. Learner Corpus Res.* 1, 7–24. doi: 10.1075/ijlcr.1.1.01gra

Granger, S., Dupont, M., Meunier, F., Naets, H., and Paquot, M. (2020). *The International Corpus of Learner English. Version 3*. Louvainla-Neuve: Presses Universitaires de Louvain.

Granger, S., Gilquin, G., and Meunier, F. (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Gries, S. T. (2013). 50-something years of work on collocations. *Int. J. Corpus Linguist.* 18, 137–166. doi: 10.1075/ijcl.18.1.09gri

Gries, S. T. (2014). "Quantitative corpus approaches to linguistic analysis: seven or eight levels of resolution and the lessons they teach us," in *Developments in English: Expanding Electronic Evidence*, eds I. Taavitsainen, M. Kytö, C. Claridge and J. Smith (Cambridge: Cambridge University Press), 29–47.

Gries, S. T. (2019). 15 years of collostructions. some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *Int. J. Corpus Linguist.* 24, 385–412. doi: 10.1075/ijcl.00011.gri

Gries, S. T., and Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: two suggestions. *Corpora* 9, 109–136. doi: 10.3366/cor.2014.0053

Gries, S. T., and Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Lang. Learn.* 65, 228–255. doi: 10.1111/lang.12119

Gries, S. T., and Kootstra, G. J. (2017). Structural priming within and across languages: a corpus-based perspective. *Lang. Cogn.* 20, 235–250. doi: 10.1017/S1366728916001085

Gries, S. T., and Wulff, S. (2005). Do foreign language learners also have constructions? *Ann. Rev. Cogn. Linguist.* 3, 182–200. doi: 10.1075/arcl.3.10gri

Gurzynski-Weiss, L., Geeslin, K. L., Daidone, D., Linford, B., Long, A. Y., Michalski, I., et al. (2018). "Examining multifaceted sources of input: variationist and usage-based approaches to understanding the L2 classroom," in *Usage-inspired L2 Instruction* (Amsterdam: John Benjamins), 291–311.

Hilles, S. (1991). "Access to universal grammar in second language acquisition," in *Point Counterpoint: Universal Grammar in the Second Language*, ed L. Eubank (Amsterdam: John Benjamins), 305–338.

Hilpert, M. (2017). "Frequencies in diachronic corpora and knowledge of language," in *The Changing English Language—Psycholinguistic Perspectives*, eds M. Hundt, S. Mollin and S. E. Pfenninger (Cambridge: Cambridge University Press), 49–68.

Hilpert, M. (2019). Higher-order schemas in morphology: What they are, how they work, and where to find them. *Word Struct.* 12, 261–273. doi: 10.3366/word.2019.0149

Hirschmann, H. (2015). *Modifikatoren im Deutschen: Ihre Klassifizierung und varietätenspezifische Verwendung*. Tübingen: Stauffenburg-Verlag.

Hirschmann, H., Lüdeling, A., Rehbein, I., Reznicek, M., and Zeldes, A. (2013). "Underuse of syntactic categories in Falko. A case study on modification," in *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead*, eds S. Granger and G. Gilquin (Louvain: Press Universitaire de Louvain), 223–234.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-native Speakers: Theory and Research, Volume 41 of Language Learning Language Teaching*. Amsterdam: John Benjamins.

Hulstijn, J. H. (2019). An individual-differences framework for comparing nonnative with native speakers: perspectives from BLC theory. *Lang. Learn.* 69, 157–183. doi: 10.1111/lang.12317

Jaeger, F. T., and Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127, 57–83. doi: 10.1016/j.cognition.2012.10.013

Jolsavi, H., McCauley, S. M., and Christiansen, M. H. (2013). "Meaning overrides frequency in idiomatic and compositional multiword chunks," in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz and I. Wachsmuth (Austin, TX: Cognitive Science Society), 692–697.

Kidd, E. (2012). Individual differences in syntactic priming in language acquisition. *Appl. Psycholinguist.* 33, 393–418. doi: 10.1017/S0142716411000415

Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends Cogn. Sci.* 22, 154–169. doi: 10.1016/j.tics.2017.11.006

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguist. Linguist. Theory* 1, 263–276. doi: 10.1515/cllt.2005.1.2.263

Klein, W. (1998). The contribution of second language acquisition research. *Lang. Learn.* 48, 527–549. doi: 10.1111/0023-8333.00057

Koplenig, A. (2017). Against statistical significance testing in corpus linguistics. *Corpus Linguist. Linguist. Theory* 15, 321–346. doi: 10.1515/cllt-2016-0036

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites, Vol. 1*. Stanford: Stanford University Press.

Larsen-Freeman, D. (2006). Second language acquisition and the issue of fossilization: there is no end, and there is no state. *Stud. Fossilizat. Second Lang. Acquisit.* 2005, 189–200. doi: 10.21832/9781853598371-012

Linford, B., Long, A., Solon, M., Geeslin, K., Ortega, L., Tyler, A., et al. (2016). "Measuring lexical frequency: Comparison groups and subject expression in L2 Spanish," in *The Usage-Based Study of Language Learning and Multilingualism*, (Washington, DC: Georgetown University Press), 137–154.

Lowie, W. M., and Verspoor, M. H. (2019). Individual differences and the ergodicity problem. *Lang. Learn.* 69, 184–206. doi: 10.1111/lang.12324

Lüdeling, A. (2017). *Variationistische Korpusstudien*. Berlin: De Gruyter.

Lüdeling, A., Hirschmann, H., and Shadrova, A. (2017). Linguistic models, acquisition theories, and learner corpora: morphological productivity in SLA research exemplified by complex verbs in German. *Lang. Learn.* 67, 96–129. doi: 10.1111/lang.12231

Lüdeling, A., Hirschmann, H., Shadrova, A., and Wan, S. (2021). "Tiefe Analyse von Lernerkorpora," in *Deutsch in Europa. Sprachpolitisch, Grammatisch, Methodisch*, eds H. Lobin, A. Witt and A. Wöllstein (Berlin: De Gruyter Mouton), 235–283.

Lukassek, J., Akbari, R., and Lüdeling, A. (2021). *Richtlinie zur morphologischen Annotation von Nomina in Falko*. Technical report, Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin.

Mesarovic, M. D. (1964). "Foundations for a general systems theory," in *Proceedings of the Second Systems Symposium at Case Institute of Technology: Views on General Systems Theory* (New York, NY: John Wiley & Sons), 1–24.

Mulder, K., and Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Appl. Linguist.* 32, 475–494. doi: 10.1093/applin/amr016

Nivre, J., Hall, J., and Nilsson, J. (2006). "Maltparser: a data-driven parser-generator for dependency parsing," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, *Vol. 6*, (Genoa: European Language Resources Association), 2216–2219.

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

Ortega, L. (2015a). "Second language learning explained? SLA across 10 contemporary theories," in *Theories in Second Language Acquisition - An Introduction, Second Language Acquisition Research Series, Chapter 5*, eds B. Van Patten and J. Williams (New York,NY: London: Routledge), 245–272.

Ortega, L. (2015b). "Usage-based SLA: a research habitus whose time has come," in *Usage-Based Perspectives on Second Language Learning*, eds T. Cadierno and S. W. Eskildsen (Berlin: De Gruyter Mouton), 353–374.

Paquot, M., and Granger, S. (2012). Formulaic language in learner corpora. *Annu. Rev. Appl. Linguist.* 32, 130–149. doi: 10.1017/S0267190512000098

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: a critical review and future directions. *Psychon. Bull. Rev.* 21, 1112–1130. doi: 10.3758/s13423-014-0585-6

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reznicek, M., Lüdeling, A., and Hirschmann, H. (2013). "Competing target hypotheses in the Falko corpus," in *Automatic Treatment and Analysis of Learner Corpus Data, volume 59 of Studies in Corpus Linguistics*, eds A. Díaz-Negrillo, N. Ballier and P. Thompson (Amsterdam: John Benjamins), 101–123.

Reznicek, M., Walter, M., Schmidt, K., Lüdeling, A., Hirschmann, H., Krummes, C., et al. (2012). *Das Falko-Handbuch: Korpusaufbau und Annotationen*. Technical report, Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin.

Roelofs, A. (2008). Dynamics of the attentional control of word retrieval: analyses of response time distributions. *J. Exp. Psychol. Gen.* 137, 303–323. doi: 10.1037/0096-3445.137.2.303

Rothman, J., and Iverson, M. (2008). Poverty-of-the-stimulus and SLA epistemology: considering L2 knowledge of aspectual phrasal semantics. *Lang. Acquis.* 15, 270–314. doi: 10.1080/10489220802352206

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc.

Ruth, L., and Murphy, S. (1988). *Designing Writing Tasks for the Assessment of Writing. Perspectives in Writing Research*. Norwood: Ablex Publishing Corporation.

Sag, I. A. (2012). "Sign-based construction grammar: an informal synopsis," in *Sign-based Construction Grammar*, eds H. C. Boasand I. A. Sag (Stanford, CA: CSLI Publications), 69–202.

Saussure, F. D. ([1916] 1983). *Course in General Linguistics*, Transl. by Roy Harris. London: Duckworth.

Schiller, A., Teufel, S., and Thielen, C. (1995). *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Technical report, Universities of Stuttgart and Tübingen.

Schmid, H. (1994). Part-of-speech tagging with neural networks. *CoRR, abs/cmp-lg/9410018*. doi: 10.3115/991886.991915

Schmid, H.-J. (2010). "Does frequency in text instantiate entrenchment in the cognitive system," in *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, eds D. Glynn and K. Fischer (Berlin: de Gruyter). 101–133.

Schmitt, N. (ed.). (2004). *Formulaic Sequences: Acquisition, Processing and Use, volume 9 of Language Learning Language Teaching*. Amsterdam: John Benjamins.

Seton, B., and Schmid, M. (2016). "Multi-competence and first language attrition," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 338–354.

Shadrova, A. (2020). *Measuring Coselectional Constraint in Learner Corpora: A Graph-based Approach* (Ph.D. thesis). Berlin: Humboldt-Universität zu Berlin.

Shadrova, A. (2020). "It may be in the structure, not the combinations: Graph metrics as an alternative to statistical measures in corpus-linguistic research," in *Proceedings of Graph Technologies in the Humanities*, eds A. Kuczera and F. Diehr.

Skehan, P. (1989). *Individual Differences in Second Language Learning*. London: Arnold.

Stefanowitsch, A., and Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *Int. J. Corpus Linguist*. 8, 209–243. doi: 10.1075/ijcl.8.2.03ste

Szmrecsanyi, B. (2005). Language users as creatures of habit: a corpus-based analysis of persistence in spoken English. *Corpus Linguist. Linguist. Theory* 1, 113–150. doi: 10.1515/cllt.2005.1.1.113

Szmrecsanyi, B. (2017). Variationist sociolinguistics and corpus-based variationist linguistics: overlap and cross-pollination potential. *Can. J. Linguist*. 62, 685–701. doi: 10.1017/cnj.2017.34

Szmrecsanyi, B. (2019). Register in variationist linguistics. *Register Stud*. 1, 76–99. doi: 10.1075/rs.18006.szm

Szmrescanyi, B. (2006). *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis., volume 177 of Trends in Linguistics. Studies and Monographs*. Berlin: De Gruyter Mouton.

Teich, E. (2003). *Cross-Linguistic Variation in System and Text. A Methodology for the Investigation of Translation and Comparable Texts, volume 5 of* Text, Translation, Computational Processing. Berlin: De Gruyter Mouton.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends Cogn. Sci*. 4, 156–163. doi: 10.1016/S1364-6613(00)01462-5

Tomasello, M. (2009). *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge, USA: Harvard University Press.

Wan, S. (2021). *Argumentationsstrategien von chinesischen Deutschlernern*. Univ.diss., Humboldt-Universität zu Berlin.

White, L. (2015). "Linguistic theory, universal grammar, and second language acquisition," in *Theories in Second Language Acquisition - An Introduction, Second Language Acquisition Research Series, Chapter 3, 2nd Edn*, eds B. Van Patten and J. Williams (New York, NY; London: Routledge), 34–53.

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

Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Softw*. 21, 1–20. doi: 10.18637/jss.v021.i12

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Wickham, H., François, R., Henry, L., and Müller, K. (2018). dplyr: a grammar of data manipulation. *R package version 0.7.6*.

Wiese, H. (2020). "Language situations: a method for capturing variation within speakers' repertoires," in *Methods in Dialectology XVI, volume 59 of Bamberg Studies in English Linguistics*, ed Y. Asahi (Frankfurt a. M: Peter Lang), 105–117.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends Cogn. Sci*. 8, 451–456. doi: 10.1016/j.tics.2004.08.006

Yoon, H.-J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System* 66,130–141. doi: 10.1016/j.system.2017.03.007

Zeldes, A. (2012). *Productivity in Argument Selection: From Morphology to Syntax, volume 260 of Trends in Linguistics. Studies and Monographs*. Berlin: De Gruyter Mouton.

Zeldes, A. (2013). Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch. *Zeitschrift für germanistische Linguistik* 41, 240–276. doi: 10.1515/zgl-2013-0014

Zinsmeister, H., Reznicek, M., Brede, J. R., Rosén, C., and Skiba, D. (2012). Das wissenschaftliche Netzwerk "Kobalt-DaF". *Zeitschrift für germanistische Linguistik* 40, 457–458. doi: 10.1515/zgl-2012-0030

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

frontiers in Communication

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fcomm.2021.737017

Individual Sensitivity to Change in the Lingua Franca Use of English

Irene Taipale[1]* *and Mikko Laitinen*[1,2]

[1]Department of Foreign Languages and Translation Studies, University of Eastern Finland, Joensuu, Finland, [2]Center for Data Intensive Sciences and Applications, Linnaeus University, Växjö, Sweden

The study of ongoing change in English typically focuses on studying evidence from codified varieties of the language. Recent corpus studies show, however, that advanced non-native users of English may display heightened sensitivity to features undergoing frequency shifts similar to that experienced by native speakers. In addition, most studies aiming to detect patterns of linguistic regularity utilize large data sets that attempt to minimize the presence of the individual. In this study, we focus on change in ELF and place non-native individuals at the center of attention. Our empirical section examines how aggregated features that are currently undergoing change in codified varieties of English vary in the repertoires of ELF users of Twitter. To carry out this task, this study utilizes geo-tagged tweets retrieved from the Nordic Tweet Stream. The data obtained from this real-time monitor corpus are freely available for research and re-use at https://cs.uef.fi/nts/. For the analysis itself, we selected the idiolects of 150 individual users who actively tweet in English from geographically varying locations in Finland. As American English predominates with several patterns of linguistic change in codified varieties of English, a simplified dichotomy between American and British features is utilized as a conceptual tool for inspecting variation. The idiolects are analyzed from the perspective of spelling and lexico-grammatical and morphological variation, such as V + *-ing* |V + infinitive (e.g. *start doing* | *start to do*) and expanded predicates (e.g. *take a look* | *have a look*). The quantitative observations show that, particularly in the case of grammatical features, ELF speakers appear to have generally adhered to ongoing linguistic change.

Keywords: individual variation, idiolect, ongoing change, English as a lingua franca, second language, sensitivity, americanization

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez, University of the Balearic Islands, Spain

Reviewed by:
Al Ryanne Gatcho, International Islamic University Malaysia, Malaysia
Ali Zangoei, University of Gonabad, Iran

***Correspondence:**
Irene Taipale
irene.taipale@uef.fi

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

Received: *06 July 2021*
Accepted: *08 December 2021*
Published: *27 January 2022*

Citation:
Taipale I and Laitinen M (2022) Individual Sensitivity to Change in the Lingua Franca Use of English. Front. Commun. 6:737017. doi: 10.3389/fcomm.2021.737017

INTRODUCTION

This study focuses on ongoing frequency shifts in English as observed in advanced lingua franca use of English (ELF). The study of ongoing shifts has mainly focused on evidence from codified and standard varieties of English and has, in consequence, concentrated on the role played by native speakers as the driving force of change. However, digitalization has brought about a situation in which a very large number of people nowadays use English as a common resource in naturalistic settings in social media (Leetaru et al., 2013; Gonçalves et al., 2018). As an illustration, on some platforms the share of English social media posts in the Nordic region covers roughly 30–40% of all the available data (Coats 2019). This has created a unique sociolinguistic setting, which calls for novel empirical approaches in studying the role of advanced non-native users of English. What is more, it also opens up theoretical approaches to rethink the role played by multilingual individuals in language change.

As an illustration of some of the new theoretical openings, recent sociolinguistic evidence suggests that advanced ELF usage in social media is embedded in social settings in a way that has been understudied previously and calls for further investigation of the role played by non-native individuals. In a computational sociolinguistic investigation, Laitinen and Lundberg (2020) use metadata from tens of thousands of social media accounts from the Nordic region, notably involving a dataset that has been automatically pruned to exclude public figures and celebrities. Their study focuses on the size and quality of the social networks of multilingual individuals using Twitter. The results show that the social networks of those who prefer English over national/domestic languages tend to be substantially different from those of other groups. Their networks are larger and comprise a substantially greater number of weak ties, suggesting that the social embedding of advanced ELF use results in sociolinguistic conditions that promote change. According to a widely accepted view in sociolinguistics, weak tie conditions in which individuals are predominantly linked through occasional and insignificant ties (such as acquaintances) tend to support innovation diffusion (Milroy and Milroy 1978; Milroy and Milroy, 1985; Milroy, 1987; Lippi-Green 1989). The underlying reason for the strength of the weak ties is that adopting new ideas is always socially risky, while there is a smaller risk of disruptive outcomes in loose-knit settings than in close-knit surroundings.

Despite its being an observation taken from a single study, this finding shifts the focus to the level of individuals and calls for further research into the role played by multilingual and multicompetent non-native individuals in language change. Sociolinguistic embedding makes ELF a unique testbed for observing ongoing frequency shifts. Weak ties and the propensity for change in general could explain why advanced ELF users may display heightened sensitivity to features that are undergoing frequency shifts (Laitinen 2016). Corpus data suggest that the use of ELF does not follow the pattern observed elsewhere whereby post-colonial second language varieties in many contexts prove to be more conservative than native varieties (Collins 2009). Rather, the quantitative patterns observed in ELF are more complex, with ELF often trailing the leading variety; it may also be that recent frequency shifts are further polarized by ELF users.

We operationalize ongoing frequency shifts through Americanization, which in previous studies has been considered to be one of the main forces of change in present-day English (Leech et al., 2009; Baker, 2017). In turn, this points to processes whereby those forms that are more frequently found in American English (AmE) tend to spread to other varieties of English, and closely overlaps with other processes such as colloquialization and informalization. Partly because of this, it is also difficult to determine whether these trends of change can be rightfully described as Americanization, or if these changes simultaneously take place in many Englishes independently of one another. The dichotomy perceived between American and British features is naturally a simplification, and national preferences refer primarily to quantitative tendencies observable in large datasets rather than strict opposites. In addition, the putative distinction also suggests that American English is a single entity, which is a partial misconception since it is subject to change and variation itself, as shown, for example, by Grieve (2016). The term ought therefore to be regarded simply as a conceptual tool that facilitates accessing ongoing frequency shifts and observing them in ELF data. Americanization makes it possible to pool a set of aggregate linguistic variables, thus and enabling researchers to distance themselves from excessive use of randomly selected individual variables, concomitantly increasing the empirical validity of their observations.

In addition to shifting the focus to the individual level, this investigation extends to two less frequently studied areas of non-native speaker sensitivity to change. Firstly, it uses geo-tagged tweets that enable zooming in to the individual level with precision. Our observations are drawn from the *Nordic Tweet Stream*, which is a real-time monitor corpus that is freely available for research and re-use at https://cs.uef.fi/nts/ (Laitinen et al., 2018). As we show in the following section, some previous studies of Americanization have used textual data from Twitter as their material. For instance, the global survey of Americanization in Englishes around the world by Gonçalves et al. (2018: 8) made use of a large collection (c. 31mi) of tweets, but despite this they show how "the real problem is the lack of data in certain cells" from countries in which English is not the main language. The NTS has been collected to fix the small data problem in such ELF settings. Secondly, most studies aiming at detecting patterns of linguistic regularity utilize datasets that either attempt to minimize the presence of the individual or rely on a very small number of individuals. This study places advanced non-native individuals at the center of attention as we investigate the idiolects of 150 individual users who actively tweet in English from geographically varying locations in Finland.

This article aims to provide an answer to the following research questions:

1) To what extent and in the usage of which linguistic variables are individual, advanced Finnish ELF users sensitive to an Americanization trend detected in codified varieties of English?
2) Do individuals from geographically varying locations and varyingly large online social networks exhibit different patterns of sensitivity?

The structure of this article consists of the following. **Section 2** introduces relevant previous literature and lacunae in the study of non-native English. **Section 3** details our material and methods, and **Section 4** discusses our observations. Finally, **Section 5** discusses the implications of our findings and presents thoughts on topic areas for future sociolinguistic study of advanced non-native Englishes.

BACKGROUND AND PREVIOUS STUDIES

Frequency shifts connected to Americanization have become increasingly prevalent in recent decades (Leech et al., 2009: 252–259; Baker 2017). In Mair (2013) system of world Englishes, he suggests that it is appropriate to label AmE as the "hub variety," a variety that is relevant to speakers of all other varieties. Keys to understanding this hierarchical system of

Englishes, as proposed by Mair, can be found in social factors, and he argues that the fact that linguistic forms preferred in AmE have spread to other varieties stems from the demographic weight of the US and a range of institutional factors prevailing in the country. These factors include the political, economic and military pre-eminence of the United States.

The phenomenon of Americanization has been observed in the study of lexis and grammar in native metropolitan varieties of English (Depraetere, 2003; Leech et al., 2009; Rohdenburg and Schlüter, 2009a; Baker, 2017). In addition, various post-colonial varieties have been investigated, including Americanisms in Nigerian English phonology and lexis (Awonusi 1994), the study of spelling and lexis in Kenyan, Singaporean, and Trinidad and Tobagoan Englishes (Hänsel and Deuber 2013). A range of studies has focused on empirical evidence concerning Philippine English syntax (Schneider 2011; Collins et al., 2014a; Collins et al., 2014b; Collins 2015), while the spelling and lexis of this variety may be found in Fuchs (2017).

In terms of data, the majority of past studies rely primarily on stratified corpus data that represent standard and edited language use, but some novel approaches also use societal big data from freely-available sources, such as the APIs of various social media applications. As an illustration, a global survey of English by Gonçalves et al. (2018) makes use of geo-located tweets written in English, thus expanding the empirical basis towards more naturalistic language use from across the world and improving the empirical validity in the study of non-native production. Their results show the strong presence of American English variants across the globe, except in countries where colonial "British influence has been strongest" (2018: 12). In addition, the data show that in non-native settings in the expanding circle contexts "American orthography and vocabulary dominate" (2018: 9–10).

Past approaches that focus on Americanization across varieties have focused on linguistic analyses, while in the process they have downplayed the sociolinguistic aspects. We argue that being sensitive to change ultimately makes the quantitative study of all non-native uses a variationist sociolinguistic endeavor. According to Weinreich et al. (1968), the most essential aspect in the study of linguistic change resides in orderly heterogeneity, in which variation is not random but systematic. To understand this systematicity and to study change in social context involves a range of levels of analysis. Its actuation involves charting for the emergence of innovative linguistic items, while change can be constrained by a range of systemic factors. Transition focuses on investigating how change proceeds from one stage to another, from the individual to the community. The question of embedding examines how change is embedded in the linguistic and social structures in language use, while social evaluation takes into account how linguistic forms are assessed by users. The core is to approach forms as the linguistics variables of writers' choices between alternative ways of expressing the same meaning. By applying a variationist approach to our data, we are able to look into how frequency shifts might be embedded in the social structures and how they are evaluated.

The variationist approach offers us a systematic apparatus for investigating variation and change in ELF (see Laitinen and Lundberg 2020 for a study of embedding). Despite the obvious benefits, variationist approaches have not yet been widely used in the study of ELF, as micro-level investigations have predominantly focused on the cognitive processing capabilities of individuals (Vetchinnikova 2017; Mauranen 2018). The role played by an individual has been more extensively highlighted recently by Vetchinnikova and Hiltunen (2020), who argue for the need to observe variability in ELF on the individual level. Their study sets out to investigate the extent to which variability in advanced ELF use stems from individual variation. Observation of ten individuals in online ELF environments reveals that the individual and the communal levels are different and the communal level ought to be "seen as emergent from the individual" and as being qualitatively different from it (also Vetchinnikova 2017).

In addition to social embedding, sociolinguistic evaluation is closely linked to the setting from which we have drawn our data. Our primary data consist of English data from Finland, where English is used as an additional language alongside the national languages (Finnish and Swedish). It goes without saying that ELF in such contexts is not a focused variety with a clearly defined speech community or a widely-agreed set of norms and common structural basis. However, language users are faced with a choice related to the norms of what the preferred language variant is. Large-scale survey data reflecting attitudes in this setting suggest that American and British varieties are generally considered prestigious. In Leppänen et al. (2011: 70–74), the respondents were requested to assess which varieties of English they found most appealing. The findings show that people in urban areas prefer British English, whereas those living in the countryside prefer American English. The division is clear since the more highly educated respondents – in general, those living in urban areas and older informants – prefer BrE, whereas the respondents living in the countryside and young informants in general prefer AmE (Leppänen et al. 2011: 73).

The following section details a data source that provides researchers with access to naturalistic advanced non-native use of English on one social media platform. The dataset contains geo-location information which makes it possible to access non-native language use from multilingual populations in various regions, thus opening up new empirical avenues in the study of advanced ELF use.

MATERIAL AND METHODS

Three sets of data from geographically distinct areas were compiled using *the Nordic Tweet Stream* (NTS). The data consist of tweets from the time period of June 1, 2016 to April 30, 2017. The first set of data aims at representing the largest cities in Finland and includes tweets from Helsinki, Tampere, Oulu, Turku, and Jyväskylä. The second set represents medium-sized localities, while the third group contains data from the countryside and small towns. It consists of small amounts of data from several different locations in all parts of the country. Each regional set of data consists of the idiolects of 50 users, resulting in altogether 150 idiolects for analysis.

Taipale and Laitinen

The sorting was, for the most part, automated. However, some manual checking was executed, as the API crawler falsely identified many English tweets as instances of some other language. In addition, spam, tweets automatically generated by third-party applications, and extensive amounts of news and forecast reports written in a repetitive form were removed from the data through manual inspection.

According to recent estimates, geo-tagged tweets offer a good overview of local language ecologies despite representing only a small fraction of the material (see, for example, Lamanna et al., 2018). However, to reduce the possible effects of traveling, additional information was used to fine-tune the data. The NTS collects various kinds of metadata ranging from the time of the tweet being sent to the information that the users have written about themselves on their profile. The tweets were sorted by discarding those that were sent from a user who has clearly set their hometown set to something other than the three areas mentioned earlier. For example, the tweet was removed in cases where a tweet written in the countryside was sent by a user who has inserted Helsinki as their hometown.

The information available about the most active users was also inspected qualitatively to see if there were further signs of their nationality and place of residence. Accounts that appeared to belong to people who were clearly from countries other than Finland were also removed from the analysis. The tweets of the 50 most active users in each region have been selected for this final examination.

Subsequently, the final data set was syntactically annotated using the GATE Twitter part-of-speech tagger (Derczynski et al., 2013) designed specifically for tagging tweets. The tagger utilizes the Penn Treebank part-of-speech tags. It is approximated to achieve 91% accuracy, which we consider to be acceptable for our research questions in this text category. The final set of data consists of a total of 34,830 tweets and 546,542 tokens.

With regard to methods, we focus on variables in the spelling and in the lexico-grammar. Spelling differences among AmE and BrE are relatively stable and may be known to laymen, whereas empirical evidence suggests that most of the grammatical variables are undergoing change, which we assume the ELF users are unlikely to be aware of. Comparing non-native usage of these kinds of variables offers a unique perspective into the study of intraindividual variation. The analysis examines a total of seven spelling variables in which AmE and BrE tendencies differ significantly. The variables come from a total of 60 distinct types or word stems, which were searched for with the aid of WordSmith Tools 7. The full list of the individual words is presented in **Appendix 1**.

A methodological issue concerns the differences between American and British English. A fundamental problem ingrained in many publications addressing the topic is that they merely state that certain spelling conventions and lexical items set BrE and AmE apart, but without empirical evidence to support these claims (Trudgill and Hannah 2008). The variables selected have been chosen on the basis of a number of empirical studies concerning the differences between AmE and BrE and how they have adapted to change (Mair 2006; Vosberg 2006; Hundt 2009; Leech et al., 2009; Rohdenburg and Schlüter 2009b; Tieken-Boon van Ostade 2009: 38; Tottie 2009; Ishikawa 2011; Baker 2017).

When the variables of this study are compared to a similar study by Gonçalves et al. (2018), some differences emerge. For example, *-ize | -ise* is omitted from Gonçalves et al. (2018). Other scholars such as Baker (2017: 35) are also wary of considering *-ize | -ise* as a true spelling difference between AmE and BrE, as it has been speculated to be a by-product of word-processing software. Regardless of the origins of this prevalence of *-ise* in BrE, it seems that *-ise* is nevertheless regularly used in BrE (Tieken-Boon van Ostade 2009: 38; Ishikawa 2011), which is why it is included in this study. In addition, Gonçalves et al. (2018) examine lexical variation and exclude grammatical variables from their analysis. The present study does not include lexical variation as it is prone to many pitfalls, the most challenging of them being that many lexical items known to separate AmE and BrE have slightly different meanings, which means that they are not in true co-variation with one another. As an illustration, *biscuit* is often cited as the BrE equivalent of the American *cookie,* but *biscuit* is also used in AmE to refer to a different type of pastry (Trudgill and Hannah 2008: 90; Murphy 2018: 72). In their study, Gonçalves et al. (2018) include *cookie | biscuit* and some other lexical items whose semantic equivalence is debatable (see, for example, Baker 2017: 127 on *truck | lorry*). In addition, Baker's (2017: 149) empirical findings show that lexical variation between the two varieties is rather low.

The set of aggregated spelling variables and examples of them are assembled in **Table 1**. The spelling variables selected consist of *-or | -our, -er | -re*, geminate consonants*, -ense |-ence, -yze | -yse* and some instances of *-ize | -ise* (see Appendix for a complete list of the inspected variants). In addition, some individual words have also been inspected.

The raw findings were manually pruned to exclude false positives. The example in (1) illustrates a case that was excluded from the final results. The writer displays mixed use of the AmE *gray* and BrE *grey* at the level of a single tweet, which is an accompanying note for a sepia-toned photo of a hamburger.

1) Burger in 50 Shades of Gray #Burger #50shadesgrey #Joytrip #Tampere #Koskipuisto [LC30]

One can only speculate whether this choice is conscious or not. The writer may have made a deliberate choice of using both the AmE variant *gray* in their main tweet but also the correct name of the erotic novel franchise in question, which adheres to the British norm in their hashtag. However, the hashtag is most likely not written in an attempt to entice fans of *50 Shades of Grey*, because the tweet is merely a pun and its theme is not the franchise per se. The hashtag has also been written without the preposition *of,* so even with the BrE *grey,* the formation is not the name of the novel in its totality.

Each illustration is replicated in the exact form that it appears in the original dataset, apart from the attached URLs and emoticons. Each user is given a random number based on the location of the tweet in an abbreviated form: CS is short for countryside, MST for medium-sized towns and LC for large cities. More detailed metadata, such as the exact location, has been

TABLE 1 | The examined spelling variables.

Variable	AmE example	BrE example
or | *-our*	color	colour
-er |*-re*	center	centre
geminate consonants *-l-* | *-ll-* and *-p-* | *-pp*	traveled; skillful; worshiped	travelled; skilful; worshipped
-ense | *-ence* in nouns	license	licence
-yze | *-yse* and some cases of *-ize* | *-ise*	analyze	analyse
	realize	realise
individual word spellings	toward, gray, math	towards, grey, maths

TABLE 2 | The grammatical variables used in the study.

Variable	AmE example	BrE example
gerund complement | *to*-infinitive	start doing	start to do
preterite and past participle forms of verbs such as *burn* + *get* and *have gotten/have got*	burned; have gotten	burnt; have got
downtoners	kind of, kinda	sort of, sorta
expanded predicates	take a look	have a look

excluded because the data is not analyzed at the level of individual cities, but with the privacy of the users in mind as well.

Most of the grammatical variables that are included in this study consist of lexico-grammatical items. They exist on the border between grammar and lexis, such as the variation between gerunds and *to*-infinitives in the complementation of catenative verbs that express the onset or continuation of activities or states of being (*start, begin* and *continue*). Other lexico-grammatical items include expanded predicates (e.g., *take* | *have* + *a look* etc.), the downtoners *kind of* and *sort of* and their reduced forms *kinda* | *sorta*. Some variables are morphological, such as two forms of irregular verb inflections (*burned* | *burnt*), in which the AmE variant is the more advanced. However, in the case of *gotten* | *got*, the opposite applies, which illustrates how conceptualizing AmE as the more advanced variety is a simplification as well (see Hundt 2009 for a detailed inspection of the formation of *gotten* in AmE). **Table 2** presents the chosen lexico-grammatical and morphological features.

As with the spelling, each word or word stem was searched individually. This involved manually checking the concordance lines to ensure that the instances represent relevant language uses. Firstly, spellings that are known to have separate meanings were excluded. For instance, spellings that refer to measuring devices, such as *barometer* and *thermometer,* were excluded from the analysis of *-er* | *-re*, as the spelling is the same in both AmE and BrE. In addition, many proper nouns, such as usernames, were excluded. Hashtags are included in the analysis for the most part: in cases such as *#favorite*, the hashtag does not seem to link the tweet to any specific online phenomenon. Hashtags that are most likely related to a specific chain of tweets or a phenomenon, such as *#cannabislicense,* or hashtags that are otherwise written in a fixed way, such as *#votelabour,* would have been excluded, but these kinds of hashtags were not detected in the manual analysis.

Ensuring that the hits for grammatical queries were in fact instances of the grammatical variation in question required equally careful manual pruning. For example, only some meanings of *have got* are in co-variation with *have gotten* (e.g., Biber et al., 1999: 398–400; Algeo 2006: 14), and syntactic annotation does not account for these shades of semantics. In fact, most of the initial hits were instances that referred to possession or necessity and obligation, which are not expressed via *gotten*. These kinds of false positives were excluded via a manual inspection. Similarly, search queries for expanded predicates generate thousands of false positives, as the syntactic annotator does not differentiate between expanded predicates and other verb + noun constructions, such as *take a photo*. In consequence, it was decided to search for a selection of common expanded predicates. The object nouns selected are *bath, break, look, nap* and *shower.* In the case of downtoners, instances where the downtoner precedes a noun phrase, as in (2), were excluded. This is because only grammaticalized uses of downtoners as estimators (3) appear to hold regional contrasts (Algeo 2006; Rohdenburg and Schlüter 2009b):

2) We're asking for innovations, but what kind of innovations? Pokemon games? A marvelous innovation #basicincome is not implemented, Why? [LC45]
3) I'm kind of concerned about that dog's wellbeing. XD Maybe it's time to change the owner? Based on that scream :P [MST11]

To acquire quantifiable results, we utilize an index of Americanization, adopted from Gonçalves et al. (2018). This index determines the share of variants associated in the data with American English. It also ensures comparability between variables, which vary in token frequency and provide concrete numbers to analyse the results with. This has been used for the analysis in each of its levels, ranging from individual users' preferences to normalized total data sets. It has been calculated using the following formula:

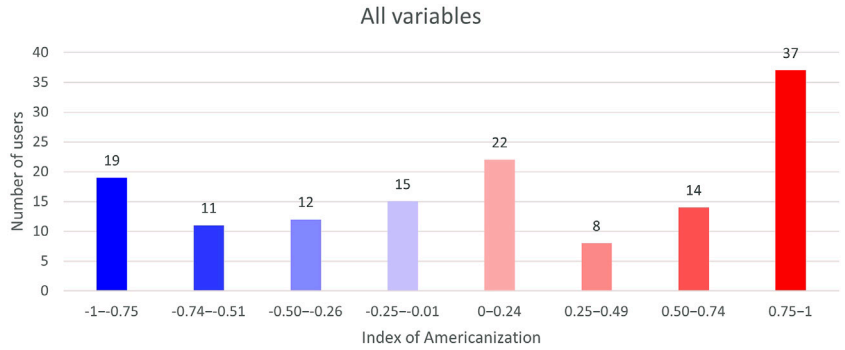

FIGURE 1 | All users' indexes of Americanization involving all variables (138 users).

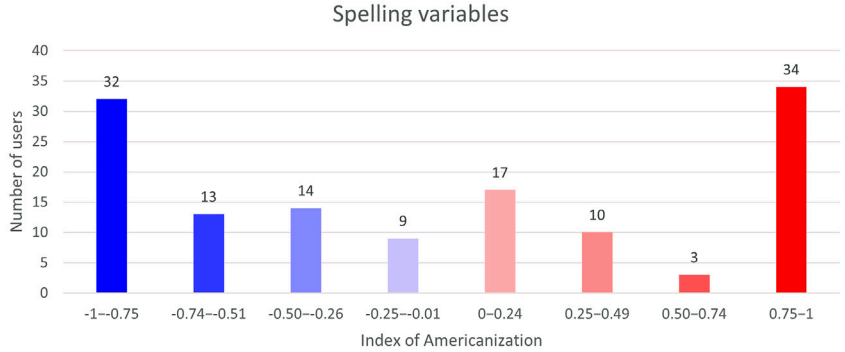

FIGURE 2 | All users' indexes of Americanization involving all spelling choices (132 users).

$$AmE\ index = \frac{(AmE\ variants - BrE\ variants)}{(AmE\ variants + BrE\ variants)}$$

Using this index, the results range from -1 to 1: an index of 1 would signify that instances of a certain variety feature conform solely to the American norm. Conversely, -1 would mean that only British variants are used. An index of 0.5 can be considered to be the threshold of highly Americanized usage.

The results are visualized with a corresponding color continuum, where an index of 1 is represented with red, -1 with blue and 0 with white; in other words, the stronger the index of Americanization, the darker the shade of red.

RESULTS

Overall Results Concerning Spelling and Lexico-grammar and Morphology

Of the 150 users' idiolects that were investigated, 138 provided relevant hits. **Figure 1** provides an overview of all users' total indexes of Americanization, meaning that all of the instances of both spelling variables and lexico-grammatical and morphological variables were examined together:

The figure displays a peak at the Americanized end of the scale: a total of 37 users appear to operate vigorously with American variants. However, a total of 57 users' total indexes land somewhere between -0.5 and 0.49, meaning that approximately 41% of all users steadily alternate between AmE and BrE forms.

Figure 2 illustrates the distribution when only the spelling variables are examined. When this restriction is applied, the users' preferences appear to be rather polarized.

In fact, the number of users that exhibit polarised usage is roughly the same at both ends of the spectrum. There is also, however, an approximately equal portion of users hovering between the spelling norms of both varieties.

The results involving lexico-grammatical and morphological variation exhibit a clear trend in their use of advanced American variants. A total of 60 users (equal to 58% of the users examined) have adopted the AmE variants extensively, as shown in **Figure 3**.

There are only a few exceptions, since 11 users appear to consistently favour the British norms. From the perspective of sensitivity, the grammatical variants that are prone to linguistic change are the most relevant. Namely, at least the vast majority of users are laymen who are most likely unaware of lexico-grammatical and morphological tendencies observed in present-day Englishes. The users are more likely to recognize regional spelling contrasts that are, for instance, taught in school, and thus they may consciously choose the more pleasing variant regardless of which variant they have been exposed to more frequently.

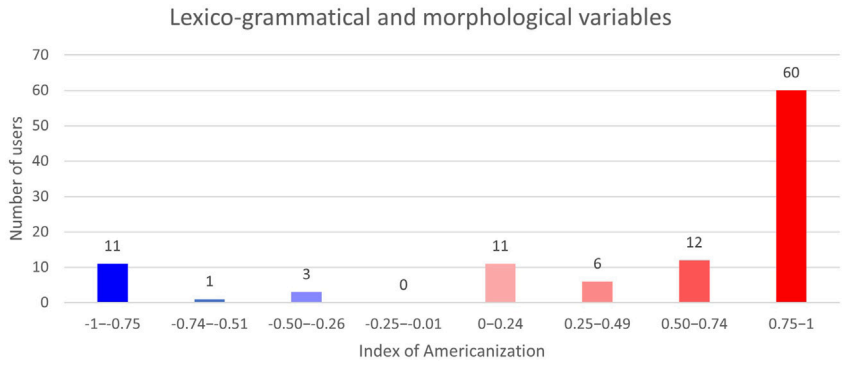

FIGURE 3 | All users' indexes of Americanization involving all grammatical features (104 users).

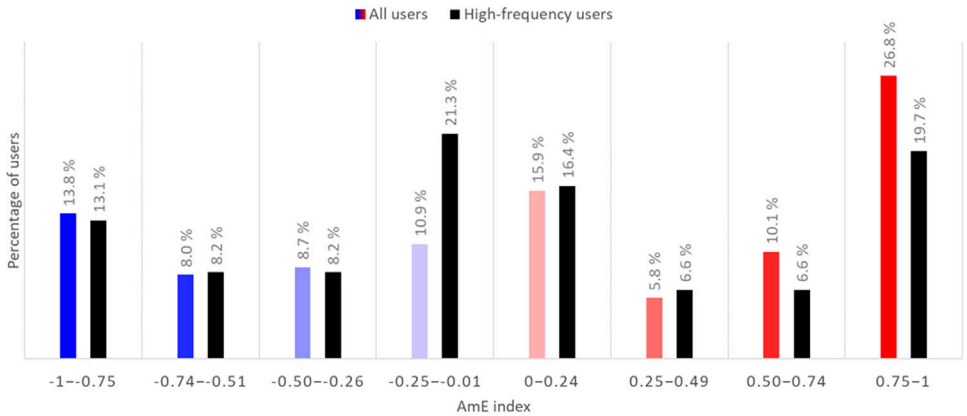

FIGURE 4 | A comparison of all users (n = 139) and users who have provided 10 or more hits (n = 61), all variables included.

However, the most polarized results must not be accepted *de facto*. Namely, many of the users who exhibit an extremely polarized usage of variants are those who have provided relevant data in lesser quantities. **Figure 4** illustrates how the repertoires of the so-called high-frequency users', i.e., those who have produced a minimum of 10 instances for inspection, differ from the data as a whole.

The two groups exhibit, for the most part, similar patterns of variation. However, mixed use of AmE and BrE variants is more common for those users who have provided 10 or more relevant linguistic occurrences. A total of 21.3% of high-frequency users' AmE indexes extend from -0.25 to -0.01, whereas when all users are examined the percentage is only 10.9. In addition, the groups that display the most Americanized use of variables differ in size, with the high-frequency users' group being 7.1% smaller.

Individual Features

When the results at the level of individual variables are inspected, several diverging patterns emerge. As a first step towards analyzing how much individual variation there is in the use of each variable, the AmE indexes of the occurrence of each variable were calculated for each user. The means of these indexes are presented **Figure 5** and **Figure 6**. **Figure 5** presents the results involving the individual spelling variables. The results have been calculated based on a total of 582 relevant instances.

As a group, individual ELF speakers appear to use the inspected spelling variables in a mixed way, the mean being close to zero (-0.06). There are also significant differences in the use of individual variables: individual BrE spellings such as *grey* and *towards* are more common than their AmE variant, whereas forms such as *-er* and *-ize*/ *-yze* lean more toward the AmE side. However, none of the spelling variables exceeds the threshold of Americanized usage (0.5); instead, the spelling variables that are closest to appearing in a form consistent with either variety are individual words that adhere to British norms with greater frequency. The polarity of this result may be explained by the fact that the variable group in question has the least individual variables. Groups of variables with roughly equal numbers of total instances (geminate consonants; *-ense* | *-ence*) appear to be used in an even more mixed way. This may indicate that some words may have a more fixed spelling than others.

However, when the users are divided into groups based on their AmE indexes and hence are examined as individuals rather

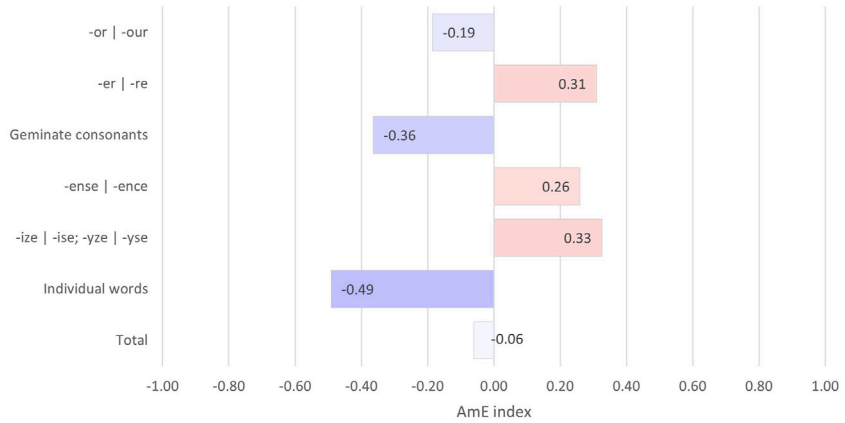

FIGURE 5 | Individual variation in the use of spelling variables presented as AmE index averages.

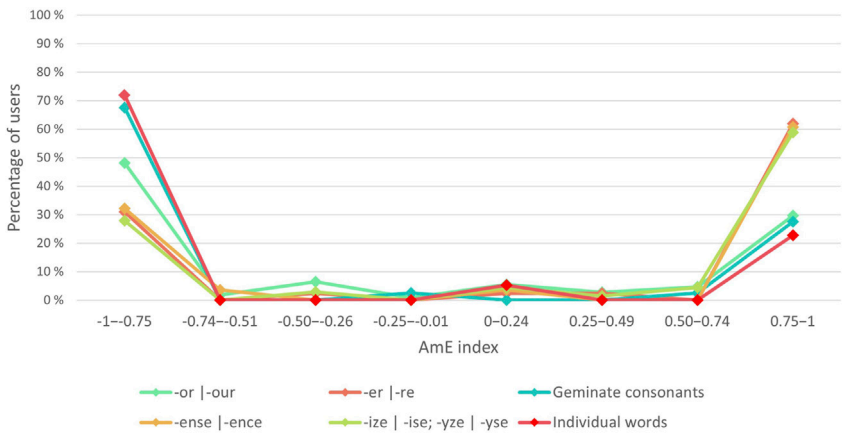

FIGURE 6 | The distribution of users' AmE indexes at the level of individual spelling variables.

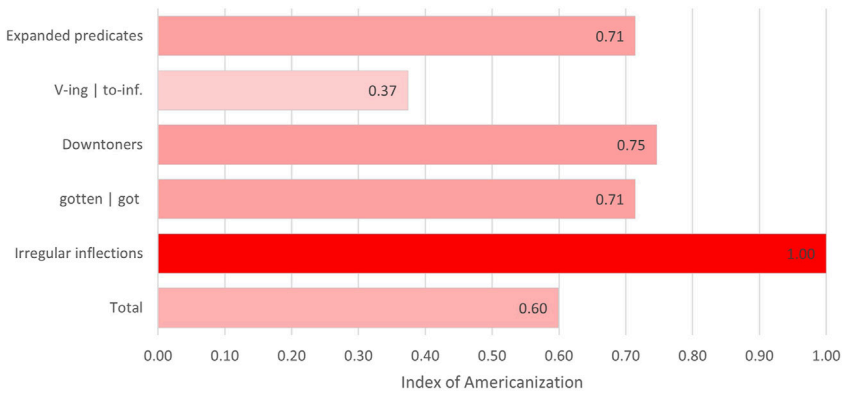

FIGURE 7 | Individual variation in the usage of morphological and lexico-grammatical variables presented as AmE index averages.

than as a group, converging patterns appear. Namely, it appears that individual users opt for either AmE or BrE variants consistently, as **Figure 6** demonstrates:

The figure shows how the majority of users are positioned at either one of the polarized ends of the spectrum, whereas mixed usage inside an individual group of variables is rare. Users often

simultaneously use AmE variants in the case of one group of words and the BrE variants in the case of another variable. The most polarized spelling variables are individual words that more than 70% of users write according to the BrE standard, while geminate consonants come second. Spelling forms that are predominantly written in the AmE style include *-er*, *-ense* and *-ize*/ *-yze*. The variation between *-or* and *-our* appears to divide users the most, as roughly half of the users strongly prefer *-our*, while approximately a third of users mostly opt for *-or*. For example, the single user MST28 in (4), (5), and 6) exemplifies a typical conjoined usage of spelling features from both metropolitan varieties (AmE *-er* and *-or* and BrE *towards*):

4) @user Off center, of course. This is such an undebatable question [MST28]
5) @user I like the EU version more, looks more colorful to me His character is fucking weird but charming. I'm weird. [MST28]
6) I feel like I'm the only person that has walked on earth who doesn't like Zelda #FlippedLearning #AI and #Robotics? Or will the pedagogic pendulum swing back towards the #lecture? [MST28]

A shift of perspective from the traditional, large-scale group and their averages to the level of individual variation provides a clear difference in outcome: what appears at the level of all users as mixed, inconsistent usage of a certain linguistic variable is in fact a result of polarities emerging from both extremities of the scale. In other words, as a larger population, Finnish ELF users use spelling variants in a mixed way, but as individuals they are usually consistent at the level of individual variables. Of course, as was noted earlier, some of the observed polarities are the result of a lack of large quantities of data.

A similar comparison of group tendencies and variation at the level of the individual was carried out in the case of lexico-grammatical and morphological variables. The averages concerning these variables are presented in **Figure 7**. The results are based on a total of 360 relevant hits.

In the case of lexico-grammatical and morphological variables, the results are more consistently Americanized than in the case of spelling, even at the level of averages. Only the variation between gerundial and infinitive complements falls below an AmE index of 0.5. Irregular inflections appear to be the most Americanized of the variables; however, a plausible reason for this result is the scarce amount of data available for this variable (a total of 19 instances), decreasing the significance of the result.

As in the observation of spelling variables above, users can be divided into groups based on their AmE indexes at the level of variables to see how consistent the individual users truly are. **Figure 8** shows that most users opt for advanced, American variants. Mixed usage of any variable is very rare, but 11% of the users have a tendency to mix *to*-infinitives and gerunds, and an even more minuscule group (5%) of users sometimes uses both *kind of/kinda* and *sort of/sorta*. In the case of verb complements, this kind of mixed usage is what could be anticipated, as the use gerunds has not been observed to completely replace *to*-infinitives (Leech et al., 2009: 195, Mair 2006: 128–130).

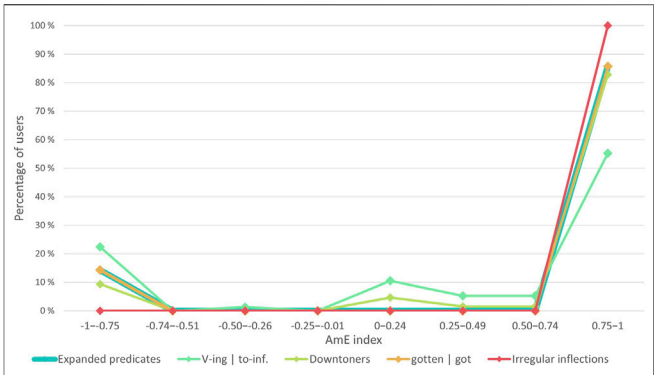

FIGURE 8 | The distribution of users' AmE indexes at the level of individual grammatical variables.

While AmE lexico-grammatical and morphological variants are used overwhelmingly, rather than BrE forms, there is room for individual variation. This is particularly evident in the usage of gerunds and *to*-infinitives, which are the least Americanized of the variables. Examples 7) and 8) show how gerunds and *to*-infinitives can appear in co-variation at the level of an idiolect. The user in question uses both the gerund and the *to*-infinitive to convey what appears to be exactly the same meaning with the same verb *hate*:

7) i lit ('literally') can't live my life without constantly being afraid that people start to hate me out of nowhere [LC32]
8) Every time when i see a goodlooking person I immediately start hating myself i can't help it akbenrkgnf [LC32]

What makes the example even more relevant is that *hate* is a stative verb, which at least according to traditional grammars usually pairs up with infinitives (e.g., Mair 2003). While there would be room for more qualitative analysis of syntactic and semantic restrictions that might govern the users' choices, examples 7) and 8) show that the users' choices are at least not fully explainable via these kinds of constraints.

Regional Variation

This section makes use of geotagging in the material and examines possible regional differences in the users' AmE indexes. To make the results more tangible, the 142 users were divided into three groups based on the level of Americanization of their idiolects. The first group, which represents those with a low AmE index, consists of individuals whose AmE index ranges from -1 to -0.51. The second group, which represents those whose repertoires are mixed, contains users with indexes between -0.5 and 0.49. The third group consists of users whose repertoires exceed the threshold of Americanized usage (0.5) and who thus may be most sensitive to ongoing change. Firstly, when all of the variables are inspected relative to the region the results take the form visualized in **Figure 9**.

As **Figure 11** illustrates, the proportion of users with a low AmE index is almost equal in every region. However, in the case of users who mix AmE and BrE variants, differences

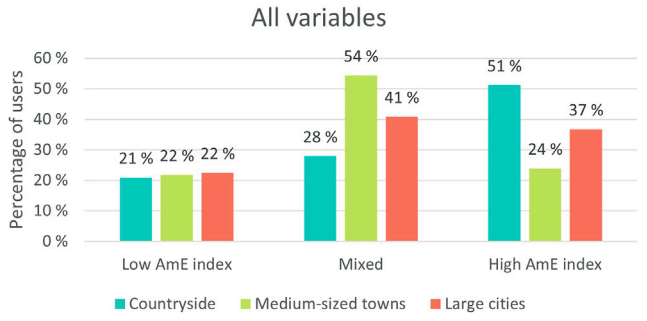

FIGURE 9 | Variation of AmE features in idiolects divided based on geographical location.

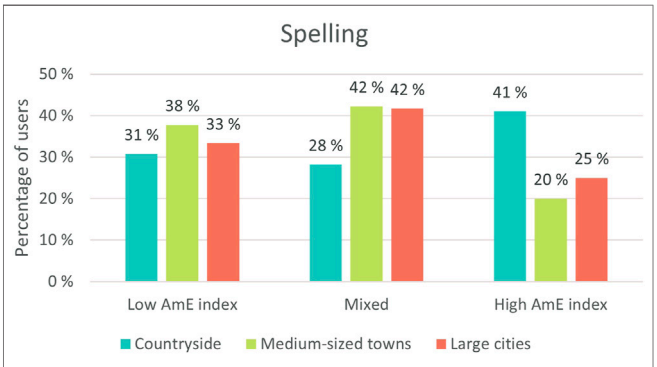

FIGURE 10 | Spelling variation in idiolects divided based on geographical location.

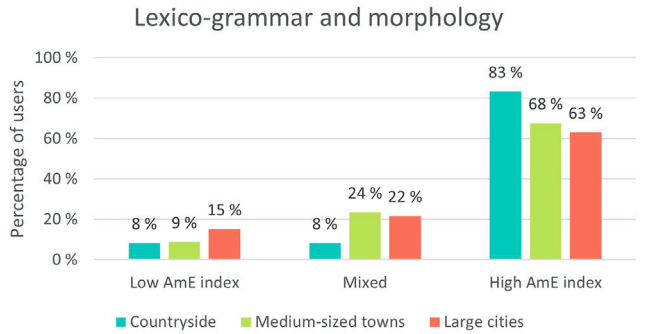

FIGURE 11 | Lexico-grammatical and morphological variation in idiolects divided based on geographical location.

emerge. Over half (54%) of the users from medium-sized towns are prone to mixing the variants. Almost a third (28%) of users from the countryside belong to this group, and an even larger number (41%) of users from large cities also hover between the AmE and BrE variants. In the case of users with a high AmE index, users from the countryside make use of American forms the most, as 51% of them have a high AmE index. Thus, it appears that this small selection of country-dwellers may be most sensitive to AmE influences. Individuals from the medium-sized towns data set, on the other hand, clearly juggle between AmE and BrE variants the most (54%), with large city users coming second (41%). The share of users whose AmE indexes are low is almost equal in each regional group.

When spelling variables are inspected by separating them from the lexico-grammatical and morphological variables, the quantitative patterns shown in **Figure 10** emerge:

As in the overall results, non-native English users from the countryside appear to be the most prone to using AmE variants, as 41% of them have high AmE indexes, whereas the approximately 42% of individuals from both medium-sized towns and large cities mix these variants. In the case of medium-sized towns and large cities, the number of users who tend toward the BrE side is also higher than the number of those whose spelling tendencies are highly Americanized.

Labov's (1966: 499) well-known distinction between "change from above" and "change from below" may be seen in the users' repertoires. In Labov's dichotomy, change is categorized according to whether change takes place above or below the boundaries of conscious awareness. In conjunction with our corpus results, Finnish people living in the countryside reportedly find American English more appealing than those living in urban areas, whereas Finns from large cities prefer BrE (Leppänen et al. 2011: 70–74). In addition, BrE is quite obviously considered a prestigious variety, and the inspected users are most likely aware of the status of BrE. It may be that people living in large cities find the prestigious British variety more compatible with their identities than country-dwellers do, and it may also be that this association is reflected in their spelling choices that may play a part in the users' construction of their social media personae. Based on these findings, it can thus be speculated, on the one hand, that via their spelling choices users may wish to make use of existing associations and collective beliefs concerning AmE and BrE and the people whose native languages these varieties are.

On the other hand, in the case of grammatical change, which most likely takes place below conscious awareness, the Americanization trend is evident in each region. Some regional differences do, however, emerge. The share of users from countryside with a high AmE index surpasses that of other regional groups, as seen in **Figure 11**:

Users from large cities, on the other hand, appear to use American lexico-grammatical and morphological forms the least. Hence, not only do country-dwellers appear to be most sensitive to American influences, but large city users appear to be the least sensitive. Thus, at least from the perspective of the raw numbers, it appears that advanced ELF users from the countryside are the most sensitive to Americanization. However, this result is not statistically significant, as the seemingly towering difference in proportion is a result of there being a relatively small number of users to compare: 24 in the countryside, 36 in medium-sized towns and 46 in large cities. Thus, the usage of American lexico-grammar and morphology appears to be a phenomenon that unifies the regions: users appear to have picked up AmE lexico-grammar and morphology regardless of their location. It would have also been counterintuitive that countryside users' repertoires

TABLE 3 | Network sizes across geographical sub-corpora.

	Median of twitter followers	Median of twitter friends	Index of strong ties
Countryside	408	378.5	0.92
Medium-sized towns	474	424.5	0.90
Large cities	717	690.5	0.96

TABLE 4 | Network sizes and users' AmE indexes.

	Median of twitter followers	Median of twitter friends	Index of strong ties
High AmE index	448	317	0.71
Mixed	447	482	1.07
Low AmE index	759	813	1.07

were significantly more advanced. Furthermore, Eisenstein et al. (2014) observe how new words that are coined in Twitter spread in a regionally tied manner from city to city, despite the possibility of social media transcending these kinds of geographical boundaries.

In addition, the social network theory predicts that weak and uniplex networks are the most fruitful platforms for change, which would make users with the largest number of friends and followers most susceptible to being agents of change. **Table 3** shows that, with regard to both friends and followers, users from countryside locations have smaller networks than those from medium-sized towns and large cites. The index of strong ties indicates the relative proportion of the users' contacts who are friends compared to followers: Twitter friendship requires reciprocity and thus indicates that the users share a stronger bond. As can be seen from **Table 3**, the countryside users' social networks, while smaller, contain fewer strong ties than the networks of users from the other areas. This kind of looseness could help to explain why users from the countryside may be more sensitive to change than the rest of the users.

Similar patterns emerge when network sizes are examined in comparison with the users' AmE indexes, as shown in **Table 4**. Contrary to what would have been expected, it appears that the lower the AmE index the larger the users' networks are. However, if instead of pure size the users' networks are approached via the index of strong ties, those with the highest AmE indexes clearly have more loose networks (0.71 vs. 1.07).

All in all, it appears that the countryside users differ from the rest of the data by exhibiting more sensitivity to Americanization than the groups of users. While this is a surprising result, it may be explained by the social network theory, which predicts that those with loose networks are prone to acting as agents of change: while their networks are smaller than the networks of people living in cities, the networks are weaker in nature. However, the results should be accepted with caution. Namely, the countryside data is the weakest in the sense that there are fewer idiolects to work with and fewer hits. Thus, a considerable portion of the individuals may only appear to use AmE variants in a polarized way. The reality may be that their polarized AmE index is simply a result of them providing only one relevant hit, and more data might reveal inconsistency. Hence, more data is needed to verify this possible tendency, but these results can thus be seen as an interesting starting point for future research with greater amounts of data.

DISCUSSION AND CONCLUSION

Our study, while preliminary, takes on understudied areas of linguistics that are related to large-scale changes in the English language continuum. We hope that this study has illustrated the potential that the study of ELF idiolects possesses to reveal in-depth patterns of change and variation, as the results are strikingly different when individuals are examined rather than the group as a whole. Our main results show that individuals exhibit considerable variation in their spelling choices and lexico-grammatical and morphological tendencies. In addition, the countryside data differs from the rest of the data: it appears that Finnish users of English from non-urban areas are perhaps most sensitive to Americanization.

With regard to our first research question, we can observe that individuals exhibit considerable variation in their respective usage of different linguistic variables. Overall, spelling tendencies appear to be mixed, whereas lexico-grammatical and morphological variation clearly tends toward AmE. Our observations were of patterns similar to those found by Gonçalves et al. (2018), but our study provides more detail on how both idiolects and individual users' usage of various spelling features varies. When analyzed as a large group, the users appear to mix AmE and BrE variants and the overall index of Americanization is neutral. However, when inspected at the more sophisticated level of idiolects and individual features, it seems that users are in fact rather polarised in their choices between spelling variants. They may also be consistent in their usage of individual spelling variables (such as *-or* vs. *-our*) but some individuals may use the AmE variant in the one spelling category and the BrE variant in another. Thus, our findings complement the macro-level observations of Gonçalves et al. (2018). By zooming in to the level of individual speakers and individual linguistic variables, we can

distinguish the role played by individual variation in the overall mixed usage of AmE and BrE variants.

With respect to our second research question, the observations show that users from the countryside may be more sensitive to change, and this may be explained by the nature of the users' social networks. By the very nature of their audience, Twitter users always communicate in a way that enables a very loose network or the general public to see their tweets. In addition, with regard to the size of the users' networks consisting of friends and followers, the users can be said to operate in loosely-knit networks. The larger the municipality the users tweet from, the larger their networks are; however, users in the countryside have the smallest number of strong ties. According to the main hypotheses of the social network theory, which suggest that diffusion takes place more readily in loose social networks, users from the largest cities were expected to be the most prone to sensitivity. However, as the proportion of strong ties is smaller in the countryside-dwellers' online social networks, countryside users can be positioned as more likely to exhibit sensitivity related to the principles of the social network theory. Overall, our findings do not confirm the hypothesis presented in the social network theory but nor do they necessarily challenge it, as the results concerning regional differences in lexico-grammatical and morphological variation are not statistically significant. Instead, it appears that users are sensitive to ongoing grammatical change in roughly equal proportions in every region.

When both spelling and grammar are included in the analysis, statistically relevant differences do, however, emerge. Thus, our findings offer empirical support for the idea that Finnish individuals' repertoires may in fact comply with their observed preferences, particularly at the level of what may be conscious decision-making. However, as noted by Baker (2017: 52), another factor that may contribute to spelling choices is computer software: for example, Microsoft Word does not make it easy for its users to deviate from the spelling conventions of the set language variety. It is unlikely that text editors are used when writing tweets, but smartphones, for example, may also favour a certain variety feature, which is something that must be taken into consideration when drawing conclusions about the results of this study. In addition, in the case of *-ize* | *-ise*, the occasional preference for *-ize* may sometimes be influenced by the preference and influence of a prestigious United Kingdom publisher, such as OUP.

All in all, our results reinforce the observation that AmE is in many respects ahead of other Englishes in the case of grammatical change and perhaps even leads these changes (cf. Leech et al., 2009; Baker 2017), and Mair's (2013) conceptualisation of Englishes can be backed up by more empirical evidence. In fact, some ELF speakers appear to be even more sensitive to change than native speakers, as lexico-grammatical and morphological choices adhere to the American forms to a strikingly high degree. Grammar may also be the layer that is more prone to unconscious choices and thus reflects language change and sensitivity more thoroughly than spelling, which users may be aware of and with regard to which they may make conscious stylistic choices. However, one should also note that present-day usage in Leech et al. (2009) is represented with corpora from 1991/2, which is almost equally distant from the actual time at which the data used in this study was written as its point of comparison, ergo corpora from the 1960s. One might expect that the pattern of change observed by Leech et al. has intensified during the 2000s. In addition, due to the absence of large amounts of data from individual users, the results of this study are in need of verification.

By making use of one of the first non-native English data sources that enable the inspection of intranational regional variation, we offer new and unique perspectives for the study of ongoing change in English. Digital microblogging data is increasingly used to complement what traditional methods of linguistic enquiry have previously brought to the field. For example, Huang et al. (2016) study American regional dialects in Twitter, and by using solid empirical evidence, their study both confirms and enriches previous understandings of US dialects. At the same time, their approach challenges the traditional ideal informant, i.e., the non-mobile, old, rural male (NORM), since Twitter data is known to over-represent the younger generations (e.g., Cramton et al., 2013; Longley et al., 2015). Utilizing data that have been produced by in many respects the polar opposite of the NORM is particularly fitting for our study, which attempts to trace ongoing change in English among a population of non-native social media users.

Regarding future research, the present study offers many potential paths for future inquiry. Firstly, the themes of the study could be approached via larger and even more sophisticated sets of data. In addition, we hope that this study could function as an inspiration for the study of other linguistic mechanisms of non-native English that were not explored in this study. This area of study would benefit from the examination of a more comprehensive set of linguistic variables by including, for example, lexical items. In addition, using up-to-date reference corpora from the native varieties would elevate the empirical reliability of future endeavors in the field. This kind of quantitative approach to idiolects is unique. In addition, the study leaves room for more qualitatively oriented viewpoints. The study of individual variation would benefit from an analysis that would include the semantic and syntactic factors that may govern the users' choices between different variants. In addition, the users' social networks could be inspected in a more fine-tuned way, such as focusing on the level of interaction between certain individuals and analyzing variation and diffusion at the level of entire networks by inspecting online social networks of varying size and type. The most fruitful path of future inquiry would naturally would naturally be one where the macro and the micro level complement one another.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://cs.uef.fi/nts/

AUTHOR CONTRIBUTIONS

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

REFERENCES

Algeo, J. (2006). *British or American English? A Handbook of Word and Grammar Patterns*. Cambridge: Cambridge University Press.

Awonusi, V. O. (1994). The Americanization of Nigerian English. *World Engl.* 13 (1), 75–82. doi:10.1111/j.1467-971x.1994.tb00284.x

Baker, P. (2017). *American and British English: Divided by a Common Language?* Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.

Coats, S. (2019). Language Choice and Gender in a Nordic Social media Corpus. *Nord. J. Linguist.* 42 (1), 31–55. doi:10.1017/s0332586519000039

Collins, P., Borlongan, A. M., Lim, J.-H., and Yao, X. (2014a). "The Subjunctive Mood in Philippine English," in *Contact, Variation, and Change in the History of English*. Editors S. Pfenninger, O. Timofeeva, A.-C. Gardner, A. Honkapohja, M. Hundt, and D. Schreier (Amsterdam: Benjamins), 259–280. doi:10.1075/slcs.159.13col

Collins, P., Borlongan, A. M., and Yao, X. (2014b). Modality in Philippine English. *J. English Linguist.* 42, 68–88. doi:10.1177/0075424213511462

Collins, P. (2009). Modals and Quasi-Modals in World Englishes. *World Engl.* 28 (3), 281–292. doi:10.1111/j.1467-971x.2009.01593.x

Collins, P. (2015). "Recent Diachronic Change in the Progressive in Philippine English," in *Grammatical Change in English World-wide*. Editor P. Collins (Amsterdam: John Benjamins), 271–296. doi:10.1075/scl.67.12col

Cramton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., and Wilson, M. (2013). Beyond the Geotag: Situating 'big Data' and Leveraging the Potential of the Geoweb. *Cartogr. Geogr. Inf. Sci.* 40 (2), 130–139. doi:10.1080/15230406.2013.777137

Depraetere, I. (2003). On Verbal concord with Collective Nouns in British English. *English Lang. Linguist.* 7 (1), 85–127. doi:10.1017/s1360674303211047

Derczynski, L., RitterClarke, A. S., and Bontcheva, K. (2013). "Twitter Part-Of-Speech Tagging for All: Overcoming Sparse and Noisy Data," in Proc. Int. Conf. Recent Adv. Nat. Lang. Process ACL, Hissar, Bulgaria, September 7-13, 2013.

Eisenstein, J., O'Connor, B., Smith, N. A., and Xing, E. P. (2014). Diffusion of Lexical Change in Social Media. *PLoS ONE* 9 (11), e113114. doi:10.1371/journal.pone.0113114

Fuchs, R. (2017). The Americanisation of Philippine English: Recent Diachronic Change in Spelling and Lexis. *Philipp. ESL J.* 19, 60–83.

Gonçalves, B., Loureiro-Porto, L., Ramasco, J. J., and Sánchez, D. (2018). Mapping the Americanization of English in Space and Time. *PLoS ONE* 13 (5), e0197741. doi:10.1371/journal.pone.0197741

Grieve, J. (2016). *Regional Variation in Written American English*. Cambridge: Cambridge University Press.

Hänsel, E. C., and Deuber, D. (2013). Globalization, Postcolonial Englishes, and the English Language Press in Kenya, Singapore, and Trinidad and Tobago. *World Engl.* 32 (3), 338–357. doi:10.1111/weng.12035

Huang, Y., Guo, D., Kasakoff, A., and Grieve, J. (2016). Understanding US regional linguistic variation with Twitter data analysis. *Comput. Environ.* 59, 244–255. doi:10.1016/j.compenvurbsys.2015.12.003

Hundt, M. (2009). "Colonial Lag, Colonial Innovation or Simply Language Change," in *One Language, Two Grammars? Differences between British and American English*. Editors G. Rohdenburg and J. Schlüter (Cambridge: Cambridge University Press), 13–37.

Ishikawa, S. (2011). Duality in the Spelling of English Verb Suffixes -ize and -ise: A Corpus-Based Study. *IPEDR* 26, 390–396.

Labov, V. (1966). *The Social Stratification of English in New York City*. New York: Harcourt. Brace.

Laitinen, M., and Lundberg, J. (2020). "ELF, Language Change, and Social Networks," in *Language Change: The Impact of English as a Lingua Franca*. Editors A. Mauranen and S. Vetchinnikova (Cambridge: Cambridge University Press), 179–204. doi:10.1017/9781108675000.011

Laitinen, M., Lundberg, J., Levin, M., and Martins, R. (2018). "The Nordic Tweet Stream: A Dynamic Real-Time Monitor Corpus of Big and Rich Language Data," in Proc. DHN 2018, Helsinki; Finland, 7 March 2018 through 9 March 2018, 349–362.

Laitinen, M. (2016). "Ongoing Changes in English Modals: On the Developments in ELF," in *New Approaches in English Linguistics: Building Bridges*. Editors O. Timofeeva, S. Chevalier, A.-C. Gardner, and A. Honkapohja (Amsterdam: John Benjamins), 175–196.

Lamanna, F., Lenormand, M., Salas-Olmedo, M. H., Romanillos, G., Gonçalves, B., and Ramasco, J. J. (2018). Immigrant Community Integration in World Cities. *PLoS One* 13 (3), e0191612. doi:10.1371/journal.pone.0191612

Leech, G., Hundt, M., Mair, C., and Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., and Shook, E. (2013). *Mapping the Global Twitter Heartbeat: The Geography of Twitter*, 18. First Monday. doi:10.5210/fm.v18i5.4366

Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., et al. (2011). National survey on the English language in Finland: Uses, meanings and attitudes. *Studies in Variation, Contacts and Change in English 5*. Available at: http://www.helsinki.fi/varieng/series/volumes/05/.

Lippi-Green, R. L. (1989). Social Network Integration and Language Change in Progress in a Rural alpine Village. *Lang. Soc.* 18, 213–234. doi:10.1017/s0047404500013476

Longley, P. A., Adnan, M., and Lansley, G. (2015). The Geotemporal Demographics of Twitter Usage. *Environ. Plan. A.* 47 (2), 465–484. doi:10.1068/a130122p

Mair, C. (2003). "Gerundial Complements after Begin and Start: Grammatical and Sociolinguistic Factors, and How They Work against Each Other," in *Determinants of Grammatical Variation in English*. Editors H. Rohdenburg and B. Mondorf (Berlin: Mouton de Gruyter), 329–345.

Mair, C. (2006). *Twentieth-Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.

Mair, C. (2013). The World System of Englishes: Accounting for the Transnational Importance of Mobile and Mediated Vernaculars. *Engl.World-wide* 34 (3), 253–278. doi:10.1075/eww.34.3.01mai

Mauranen, A. (2018). "Conceptualising ELF," in *The Handbook of English as a Lingua Franca*. Editors J. Jenkins, W. Baker, and M. Dewey (London: Routledge), 7–24.

Milroy, J., and Milroy, L. (1978). "Belfast: Change and Variation in an Urban Vernacular," in *Sociolinguistic Patterns in British English*. Editor P. Trudgill (London: Edward Arnold), 19–36.

Milroy, J., and Milroy, L. (1985). Linguistic Change, Social Network and Speaker Innovation. *J. Ling.* 21, 339–384. doi:10.1017/s0022226700010306

Milroy, L. (1987). *Language Change and Social Networks*. 2nd edition. Oxford: Blackwell.

Murphy, L. (2018). *The Prodigal Tongue: The Love-Hate Relationship between American and British English*. New York: Penguin Books.

G. Rohdenburg and J. Schlüter (Editors) (2009a). *One Language, Two Grammars? Differences between British and American English* (Cambridge: Cambridge University Press).

Rohdenburg, G., and Schlüter, J. (2009b). "New Departures," in *One Language, Two Grammars? Differences between British and American English*. Editors G. Rohdenburg and J. Schlüter (Cambridge: Cambridge University Press), 364–423.

Schneider, E. W. (2011). "The Subjunctive in Philippine English: An Updated Assessment," in *Studies of Philippine English: Exploring the Philippine Component of the International Corpus of English*. Editor M. Bautista (Manila: Anvil), 159–173.

Tieken-Boon van Ostade, I. M. (2009). *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.

Tottie, G. (2009). "How Different Are American and British English Grammar? and How Are They Different," in *One Language, Two Grammars? Differences between British and American English*. Editors G. Rohdenburg and J. Schlüter (Cambridge: Cambridge University Press), 341–363.

Trudgill, P., and Hannah, J. (2008). *International English: A Guide to Varieties of Standard English*. London: Routledge.

Vetchinnikova, S., and Hiltunen, T. (2020). "ELF and Language Change at the Individual Level," in *Language Change: The Impact of English as a Lingua Franca*. Editors A. Mauranen and S. Vetchinnikova (Cambridge: Cambridge University Press), 205–233. doi:10.1017/9781108675000.012

Vetchinnikova, S. (2017). "On the Relationship between the Cognitive and the Communal: a Complex Systems Perspective," in *Changing English: Global and Local Perspectives*. Editors M. Filppula, J. Klemola, A. Mauranen, and

Taipale and Laitinen

S. Vetchinnikova (Berlin: Mouton de Gruyter), 277–310. doi:10.1515/9783110429657-015

Vosberg, U. (2006). "The Great Complement Shift. Extra-semantic Factors Determining the Evolution of Sentential Complement Variants in Modern English," in *English and American Studies in German*, 19–22.

Weinreich, U., Labov, W., and Herzog, M. (1968). "Empirical Foundations for a Theory of Language Change," in *Directions for Historical Linguistics*. Editors W. P. Lehmann and Y. Malkiel (Austin: University of Texas Press), 95–188.

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

Taipale and Laitinen

APPENDIX 1 THE LIST OF SPELLING ITEMS (BARE FORM) USED FOR THE QUERIES.

Variable	AmE variants	BrE variants
or | -our	ardor, armor, behavior, color, colorful, endeavor, favor, flavor, harbor, honor, humor, labor, neighbor, odor, rumor, savor, tumor, vapor, vigor	ardour, armour, behaviour, colour, colourful, endeavour, favour, flavour, harbour, honour, humour, labour, neighbour, odour, rumour, savour, tumour, vapour, vigour
-er | -re	caliber, center, fiber, luster, maneuver, manoeuver, meager, meter, somber, specter, theater	calibre, centre, fibre, lustre, manoeuvre, maneuvre, manoeuvre, meagre, metre, sombre, spectre, theatre
Geminate consonants	canceled, canceling, cruelest, fulfillment, labeled, labeling, leveled, leveling, marvelous, modeled, modeling, traveled, traveling, traveler, panelist, skillful worshiped, worshiping	cancelled, cancelling, cruellest, fulfilment, labelled, labelling, levelled, levelling, marvellous, modelled, modelling, panellist, skilful, travelled, travelling, traveller, worshipped, worshipping
-ense | -ence in nouns	defense, license, offense	defence, licence, offence
-yze | -yse	analyze, analyzing, catalyze, catalyzing, dialyze, electrolyze, hydrolyze, paralyze, paralyzing	analyse, analysing, catalyse, catalysing, dialyse, electrolyse, hydrolyse, paralyse, paralysing
-ize | -ise	apologize, apologizing, criticize, criticizing, organize, organizing, realize, realizing, realization, recognize, recognizing	apologise, apologising, criticise, criticising, organise, organising, realise, realising, realisation, recognise, recognising
Individual word spellings	gray, math, toward	grey, maths, towards

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 February 2022
doi: 10.3389/fpsyg.2021.717973

Heritage Speakers as Part of the Native Language Continuum

Heike Wiese[1], Artemis Alexiadou[2,4†], Shanley Allen[5†], Oliver Bunk[1†], Natalia Gagarina[1,4†], Kateryna Iefremenko[6†], Maria Martynova[3†], Tatiana Pashkova[5†], Vicky Rizou[2†], Christoph Schroeder[6†], Anna Shadrova[1†], Luka Szucsich[3†], Rosemarie Tracy[7†], Wintai Tsehaye[7†], Sabine Zerbian[8†] and Yulia Zuban[8†]*

[1] Department of German Language and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany, [2] Department of English and American Studies, Humboldt-Universität zu Berlin, Berlin, Germany, [3] Department of Slavic and Hungarian Studies, Humboldt-Universität zu Berlin, Berlin, Germany, [4] Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany, [5] Center for Cognitive Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany, [6] Department of German, Universität Potsdam, Potsdam, Germany, [7] Department of English, Universität Mannheim, Mannheim, Germany, [8] Department of Linguistics, Universität Stuttgart, Stuttgart, Germany

OPEN ACCESS

Edited by:
Antonella Sorace, University of Edinburgh, United Kingdom

Reviewed by:
Sarah Schimke, Technical University Dortmund, Germany
Archna Bhatia, Florida Institute for Human and Machine Cognition, United States

***Correspondence:**
Heike Wiese
heike.wiese@hu-berlin.de

† These authors have contributed equally to this work

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

Received: *31 May 2021*
Accepted: *25 November 2021*
Published: *09 February 2022*

Citation:
Wiese H, Alexiadou A, Allen S, Bunk O, Gagarina N, Iefremenko K, Martynova M, Pashkova T, Rizou V, Schroeder C, Shadrova A, Szucsich L, Tracy R, Tsehaye W, Zerbian S and Zuban Y (2022) Heritage Speakers as Part of the Native Language Continuum. Front. Psychol. 12:717973. doi: 10.3389/fpsyg.2021.717973

We argue for a perspective on bilingual heritage speakers as native speakers of both their languages and present results from a large-scale, cross-linguistic study that took such a perspective and approached bilinguals and monolinguals on equal grounds. We targeted comparable language use in bilingual and monolingual speakers, crucially covering broader repertoires than just formal language. A main database was the open-access RUEG corpus, which covers comparable informal vs. formal and spoken vs. written productions by adolescent and adult bilinguals with heritage-Greek, -Russian, and -Turkish in Germany and the United States and with heritage-German in the United States, and matching data from monolinguals in Germany, the United States, Greece, Russia, and Turkey. Our main results lie in three areas. (1) We found non-canonical patterns not only in bilingual, but also in monolingual speakers, including patterns that have so far been considered absent from native grammars, in domains of morphology, syntax, intonation, and pragmatics. (2) We found a degree of lexical and morphosyntactic inter-speaker variability in monolinguals that was sometimes higher than that of bilinguals, further challenging the model of the streamlined native speaker. (3) In majority language use, non-canonical patterns were dominant in spoken and/or informal registers, and this was true for monolinguals and bilinguals. In some cases, bilingual speakers were leading quantitatively. In heritage settings where the language was not part of formal schooling, we found tendencies of register leveling, presumably due to the fact that speakers had limited access to formal registers of the heritage language. Our findings thus indicate possible quantitative differences and different register distributions rather than distinct grammatical patterns in bilingual and monolingual speakers. This supports the integration of heritage speakers into the native-speaker continuum. Approaching heritage speakers from this perspective helps us to better understand the empirical data and can shed light on language variation and change in native grammars. Furthermore, our findings for monolinguals lead us to

reconsider the state-of-the art on majority languages, given recurring evidence for non-canonical patterns that deviate from what has been assumed in the literature so far, and might have been attributed to bilingualism had we not included informal and spoken registers in monolinguals and bilinguals alike.

Keywords: heritage speakers, registers, participles, word order, bare NPs, boundary tone, referent introduction, relative clause formation

INTRODUCTION: HERITAGE SPEAKERS AS NATIVE SPEAKERS

Heritage speakers are speakers who grow up in a bi- or multilingual home with a minority language in addition to the majority language(s) dominant in the larger society (see, e.g., Montrul and Polinsky, 2011). Accordingly, these are speakers who acquire their heritage language early and naturally in a home environment as a first language, but who also acquire another language early on, which is more dominant in the larger society and will often be the only language supported in the formal context of schooling. This constitutes an interesting population that poses a challenge for the question of what counts as a "native speaker," which has traditionally been conceptualized mostly from a monolingual standpoint [cf. criticism in Ortega (2009) and Cook (2016)]. Investigating heritage speakers can help us unravel underlying assumptions of "nativeness," and contribute to our understanding of native grammars, language variation and change (e.g., Polinsky, 2015, 2018; Lowe, 2020).

The concept of "native speaker" as used in linguistic research can involve at least three different types of assumptions. First, a basic assumption is the requirement that a native speaker is "native" in the sense of being born into the language, that is, acquiring it from birth. In this sense, heritage speakers are uncontroversially native speakers of their heritage language, since they acquire it as a first language in a home environment. However, this is not the only requirement for native speakers used in the literature.

A second – implicit or explicit – requirement, often found in heritage language research, is that in order to be fully recognized as such, a native speaker has to master a repertoire that also involves standard or formal registers of their heritage language. For instance, Montrul and Polinsky (2019), while explicitly acknowledging that heritage speakers "are native speakers of their heritage language" (p. 420), require a range of registers, including formal writing, for a "proficient native speaker" (p. 426; cf. also Montrul, 2008:109). Since such registers are learned primarily in the context of formal education, heritage speakers often only acquire them for the majority language, which would then exclude them from the group of proficient native speakers of their heritage language.

It is not clear, though, why specific registers should be a necessary part of native-language proficiency, since the development of register distinctions is linked to social and communicative requirements that vary across social groups. This is independent of bi- or monolingualism, and in the case of monolinguals, it usually does not affect our view of speakers as native. For instance, we would not assume that a language without formal writing does not have any proficient native speakers, for instance in historic stages before the invention of writing, or for minority languages without a written code, and we would not claim that preschool children and people who are illiterate are not native speakers of their language.

Furthermore, register knowledge may differ substantially across (monolingual) speaker groups and registers develop throughout the lifespan, so it is not clear why some registers, but not others, would be required for native-speaker proficiency. For instance, university students will typically acquire new spoken and written registers characteristic for academia, but this will not be regarded as a requirement for being a proficient native speaker, and it is uncontroversial that monolinguals without university experience are proficient native speakers of their languages.

This suggests that we ought to keep nativeness separate from register distinctions, so that speakers who use, for instance, only informal and/or spoken registers of their language, will also be considered proficient native speakers. This underlines the importance to tap into linguistic competences supporting language use in different real-life settings, including those outside standard language (cf. also Bayram et al., 2019). Avoiding a standard language bias in our research is also important given that formal standard varieties are exceptional: they are subject to codified norms that hamper normal patterns of language variation and change, and thus should be considered peripheral rather than central instances of native language grammars.

A third kind of assumption is related to attitudinal and language-ideological patterns constructing a native speaker as someone who has grown up monolingually. Heritage speakers are primarily investigated in societies with a strong monolingual habitus (Gogolin, 1994, 2002), in particular countries based on European nation state building (including those that developed out of former European settler colonies, e.g., the United States). Against this background, a "native speaker" is often taken to be monolingual [cf. criticism in Brutt-Griffler and Samimy (2001), Bonfiglio (2010), Cook (2016), and Ortega (2016)]. This conceptualizes monolinguals as the primary owners of a language, and as the gold standard for linguistic competence and attainment. Such a conceptualization was already implicit in earlier structuralist idealizations, such as the Chomskian "ideal speaker-listener" (Chomsky, 1965) or Saussure's focus on one-to-one correlations of language and place as the "forme idéale" (de Saussure, 1916: Part 4, Ch. 2, §1). In current studies, it is implied in the use of monolinguals as a control group to test "native-like" behavior or "native competence" in heritage speakers, or to test whether some areas "develop at native levels" in heritage grammars.

However, such an idealization of the monolingual speaker as the primary bearer of a native language is not reasonable, and it is not even feasible. It is not reasonable given that multilingualism is the normal condition for human language, and, as has been amply stated (Grosjean, 1982, 2010; Romaine, 1989; Myers-Scotton, 2006, and others), most speakers in today's world are multilingual. A focus on monolinguals as native speakers in linguistics makes as little sense as the focus on males that has lately been criticized in medical research (Criado-Perez, 2019; McGregor, 2020). Medical research has, for a long time, focused only on males because studies did not want their data to be affected by hormonal changes thought to be characteristic for female bodies – but hormonal changes are part of the human condition, and if we want to know something about humans, we have to include females. In the same vein, if we want to know something about language, we have to include multilinguals, because being linguistically multi-competent is part of the human condition when it comes to language (e.g., Cook, 2016).

A restriction to monolinguals is not even feasible for empirical research, because it is not clear who would qualify as a "true monolingual." Language is always variable, speakers' repertoires always involve a range of options which could be captured as different grammars (e.g., Tracy, 2002; Roeper, 2003), and the interaction of linguistic resources within repertoires is not categorically different for languages versus dialects, registers, or styles (cf. Li Wei, 2016). This suggests that there is no clear-cut distinction between bilingual speakers who use different languages and "monolinguals" whose repertoire will always include at least different registers. Furthermore, cross-linguistic effects on the L1 have even been attested for monolingual speakers who learn a second language in an instructed, non-immersion setting (Schmid and de Leeuw, 2019). Hence, if we restrict native-speaker status to monolinguals in a strict sense, then most of the worlds' population would not count as native speakers – including most linguists today, given that most of us are fluent L2 English speakers, and thus could no longer count as native speakers of our L1.

This calls for a perspective that integrates heritage speakers into the native language continuum. Heritage language research as a field has shown that language is flexible and open to change over the lifespan. We believe that the time is ripe to take a further step and to take seriously the fact that heritage speakers are native speakers of both their languages, as emphasized in recent discussions of heritage speakers and bilingualism[1]. In what follows, we show what this means in terms of a research programme that does not take monolingual standard norms as a yardstick to identify what is missing or incorrect in heritage speakers' language use. On the basis of findings from cross-linguistic research, we show what is to be gained by overcoming a deficit-oriented view of heritage speakers. To this end, we explore the dynamics, rather than the vulnerability, of different linguistic domains and investigate development, variation, and innovation, rather than incomplete acquisition, attrition and loss. Crucially, this means (1) that we do not concentrate on standard language and formal registers alone, but capture speakers' broader repertoires, including informal and spoken language, and (2) that we target heritage speakers and monolinguals alike – not as test group vs. control group, but as two groups of native speakers that we expect to both show interesting patterns of language variation.

In what follows, we present results from a large-scale, cross-linguistic investigation that realized such a research programme within the context of the Research Unit "Emerging Grammars in Language-Contact-Situations: A Comparative View" (short "RUEG[2]")[3]. In our investigation, we approached bilinguals and monolinguals as two speaker groups to be investigated, rather than experimental vs. control group. Accordingly, we cast our net wide and targeted non-canonical patterns in general, that is, all patterns that would not be expected in standard grammar, and we did this for monolingual and bilingual groups alike.

This yielded a range of novel findings across languages, not only for heritage speakers, but also for the monolingual groups. In the following sections, we discuss evidence showing that a range of non-canonical phenomena in heritage speakers are also at work in monolingual speakers, pointing to language-internal tendencies of variation and change. These findings place multilinguals at the forefront of linguistic dynamics, and further support the integration of heritage speakers into the native language continuum. We have argued above that recognizing heritage speakers as native speakers is justified on conceptual and theoretical grounds. In what follows, we show that this perspective is also a better fit for the empirical data. We found a range of patterns that would be surprising if we saw monolinguals as a measure for "nativeness" and bilinguals as the deviant group. In contrast, these findings make a lot of sense if we see both groups as part of the native speaker continuum.

MATERIALS AND METHODS

The methods we used to elicit data meet the two demands formulated above: we need to include informal and spoken language, and we need to target non-canonical patterns and variation in bilinguals and monolinguals alike. In order to achieve this, we used the "Language Situations" ("LangSit") set-up, which avoids a restriction to formal language and taps into broader repertoires across speaker groups (cf. Wiese, 2020). In this set-up, participants are familiarized with a fictional event (e.g., a car accident) and are asked to imagine themselves as a witness to this event, and then act out telling different interlocutors about it in different communicative situations. This yields naturalistic productions that are comparable across speaker groups, languages, and settings.

[1] E.g., Rothman and Treffers-Daller (2014); Guijarro-Fuentes and Schmitz (2015), Schroeder (2016); Kupisch and Rothman (2018), Aalberse et al. (2019); Lohndal et al. (2019), Embick et al. (2020), and Flores and Rinke (2020).

[2] https://www.linguistik.hu-berlin.de/en/rueg

[3] Funded by Deutsche Forschungsgemeinschaft, DFG (FOR 2537). Speaker: Heike Wiese; further PIs: Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Mareike Keller, Anke Lüdeling, Judith Purkarthofer, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Sabine Zerbian; postdoc: Kalliopi Katsika; Ph.Ds: Katerina Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Wintai Tsehaye, and Yulia Zuban.

All materials developed for RUEG's investigation, including stimuli, elicitor instructions, and a training video for elicitors, have been stored with the Open Science Foundation for open access at https://osf.io/cm96g/.

Stimuli

For our investigation, we developed a video showing a (minor) car accident that involved a young woman with a dog, a couple with a baby in a pram, and two cars. In this video, one sees the couple approaching a car park, with the man bouncing a ball. Across a lane, the woman with the dog is unloading groceries from her car. The two cars are seen approaching the lane, when suddenly the man loses control of his ball, which bounces in front of the first car. On the other side, the dog gets excited and runs into the lane toward the ball, and the woman drops her groceries. The first car comes to an abrupt halt, causing the second one to bump into it. The man with the ball helps the woman pick up her groceries, the two drivers get out of their cars, and one of them calls the police, which can be seen through a close-up of the emergency number on his phone.

We developed five versions for five countries (see below): Germany, Greece, Russia, Turkey, and the United States. In order to support cross-linguistic comparisons, these versions only differed with respect to the emergency number, but were otherwise identical.

Procedure

Participants or their parents, in the case of adolescents, gave informed consent. For the elicitation, they saw the video, were asked to imagine themselves as a witness to the accident, and then had to play-act telling different interlocutors about it. We constructed four different communicative situations by manipulating formality and mode: participants were asked to:

(1) Leave a voice message for a friend, via instant messenger (informal-spoken).
(2) Write a message to a friend, via instant messenger (informal-written).
(3) Leave a voice message on a police "witness line" (formal-spoken).
(4) Write a witness report for the police (formal-written).

For the informal language productions (1 and 2), participants used the WhatsApp© messenger on a mobile phone provided by the elicitor, where auto correction, swiping, and suggestions had been switched off. The formal-spoken message (3) was produced on the same phone, as a voice mail to the mail box of a (fictional) contact "Police Department – eyewitness line." The formal report (4) was typed in using a simple text editor on a laptop, with spelling correction switched off.

For the different language productions, the video was shown several times. Informal versus formal productions were elicited in two different rooms that were suitably decorated according to the (in-)formality, and with two different interlocutors who acted and were dressed informally vs. formally. Short breaks filled with (in-)formal conversations divided informal and formal parts of an elicitation session.

At the end of data elicitation, participants were asked to fill in a sociolinguistic questionnaire on biographical data including language use and personality traits.

Bilingual speakers were recorded twice, in their heritage language and in the majority language, with the two sessions at least three days apart. Monolingual speakers were recorded once, in the majority language. Order of elicitation was counterbalanced for the four communicative situations, and, in the case of bilingual speakers, for the two languages.

Participants

Participants were heritage speakers and monolingual speakers. Heritage speakers were defined as speakers who had grown up with a family language in addition to the country's majority language. In order to participate, they had to use the heritage language regularly with at least some members of their nuclear family, and to be able to speak and write in it (although not necessarily in the standard alphabet). Further conditions were that they were born in the country of the respective majority language or had arrived there at an early age[4] and that they had lived in that country since, although not necessarily without interruptions. Monolingual speakers were speakers who used only one language regularly at home, namely the respective country's majority language, although they might have acquired additional languages, for instance through formal education.

The bilingual group covered heritage speakers of Greek, Russian, and Turkish in Germany and in the United States, and of German as a heritage language in the United States. The monolingual group consisted of speakers of English, German, Greek, Russian, and Turkish in the United States, Germany, Greece, Russia, and Turkey, respectively. In all categories, we covered two age groups: adolescents (14–18 years), and adults (22–35 years).

Participants had no reported speech disorders and normal or corrected-to-normal hearing and vision.

Data Processing and Corpus Generation

Elicitations yielded matched elicited, semi-spontaneous data across registers, contact-linguistic settings, and bilingual and monolingual speaker groups, in five languages:

- German as a majority language in Germany spoken by monolingual speakers, and by bilingual speakers with Greek, Russian, or Turkish as heritage languages, and as a heritage language in the United States spoken by bilingual speakers with English as a majority language;
- English as a majority language in the United States spoken by monolingual speakers, and by bilingual speakers with German, Greek, Russian, or Turkish as heritage languages;
- Greek, Russian, and Turkish as majority languages spoken by monolingual speakers in Greece, Russia, or Turkey, respectively, and as heritage languages spoken by bilingual speakers with English or German as majority languages in the United States or Germany, respectively.

[4]In general before the age of 23 months, although in some cases this was extended (up to 4 years) where otherwise it would not have been possible to recruit enough speakers.

In what follows, we refer to languages spoken as majority languages, e.g., German in Germany, or Greek in Greece, as "majority German/Greek" or short "maj-German/- Greek," etc., and to languages spoken as minority languages in a heritage context, e.g., German or Greek in the United States, as "heritage German/Greek" or short "h-German/-Greek," etc. We will use "HS" as an abbreviation for "heritage speaker." When we give examples, we provide the transcriptions for spoken data, and keep to the original spelling (including possible typos) in the case of written data.

Codes identifying data from the RUEG corpus provide the following information, in this order:

- Country: DE – Germany; GR – Greece; RU – Russia; TU – Turkey, and US – USA.
- Bi-/monolingual speaker: bi vs. mo.
- Speaker number incl. age group: 1–50 – adults; from 51 onward – adolescents.
- Gender: M vs. F (there were no speakers who identified as non-binary).
- Heritage language for bilingual speakers or only family language for monolinguals: D – German; E – English; G – Greek; R – Russian; and T – Turkish.
- Communicative situation: f – formal/i – informal and s – spoken/w – written.
- Language of production: D, E, G, R, and T.

For instance, "DEbi51MT_isD" identifies data in Germany (DE) from a bilingual (bi) adolescent (51) male (M) speaker with Turkish (T) as a h-language in an informal (i) spoken (s) setting, communicating in German (D).

All corpus data has been anonymized and integrated into a unified corpus, the RUEG corpus (Wiese et al., 2019). The RUEG corpus is a multimodal and multi-layer corpus, which in its current version (0.4.0) contains approximately 520,100 tokens[5] (appr. 146,000 for English, 157,000 for German, 66,000 for Greek, 88,000 for Russian, and 63,000 for Turkish), based on data from 716 speakers, of whom 393 are bilingual and 323 monolingual. **Table 1** gives the details for the different data sets[6].

At the time of writing, the corpus continues to grow, with more data sets and improved annotations added. Corpus data includes language productions in all four communicative situations, with additional transcriptions for spoken data (conditions 1 and 3), and the biographical data from the speaker questionnaires. Language productions are annotated for syntactic spans, lemmata, language, and parts of speech in a universal and a language-specific set of categories. The corpus can be used via the ANNIS corpus search and visualization tool (Krause, 2019). The complete corpus, including its source data and all preliminary versions, is freely available in an open repository (doi: 10.5281/zenodo.3236068).

TABLE 1 | RUEG corpus data.

Country	Bi-/monolingual	Languages	# speakers	# tokens
DE	Bilingual	maj-German	44	21,339
		h-Greek	47	19,783
	Bilingual	maj-German	56	34,503
		h-Russian	58	32,882
	Bilingual	maj-German	65	35,881
		h-Turkish	65	23,722
	Monolingual	maj-German	64	50,706
United States	Bilingual	maj-English	34	16,765
		h-German	34	14,888
	Bilingual	maj-English	64	30,913
		h-Greek	64	18,032
	Bilingual	maj-English	65	36,021
		h-Russian	66	29,214
	Bilingual	maj-English	59	32,905
		h-Turkish	56	18,502
	Monolingual	maj-English	64	29,238
GR	Monolingual	maj-Greek	64	27,931
RU	Monolingual	maj-Russian	67	25,930
TU	Monolingual	maj-Turkish	64	20,947

RESULTS

Our analyses yield three main findings: (1) cross-linguistically, we find non-canonical patterns not only in heritage speakers, but also in monolinguals, including patterns that, according to the literature, would not be expected for monolinguals; (2) we find extensive variation not only in heritage speakers, but also in monolinguals; (3) non-canonical patterns interact with register, underlining the importance of taking into account both formal and informal settings and, crucially, doing so for multilinguals and monolinguals alike.

Non-canonical Patterns: Not Just in Heritage Speakers, but Also in Monolinguals

In order to demonstrate what can be gained by approaching heritage speakers as native speakers of both their languages, we present non-canonical patterns that we observed in both heritage speakers and monolinguals. These are patterns that have so far been considered absent from native grammars and which might have been attributed to bilingualism had we taken a less inclusive approach. We cover domains of morphology and syntax, intonation, and pragmatics. In what follows, we present results from different languages, combining, in each case, qualitative and quantitative analyses. Qualitative analyses capture the relevant patterns and their distribution across bilingual and monolingual speaker groups. Quantitative analyses compare frequencies between different groups. In domains where corpus frequencies are high enough, this is supported by statistical tests[7]. For lower-frequency phenomena, we provide comparative

[5]We define "token" as the minimal annotated unit in our corpus: a string of characters between two spaces in written text, or in the written transcription of spoken text. Thus, a token is typically a word (e.g., "dog") or an emoji or emoticon (e.g.,:-/). Punctuation marks and filled pauses (e.g., "um") are not included as tokens.

[6]Note that the "bilingual" rows represent the same speakers on two lines – e.g., maj-German (44) and h-Greek (47) are the same speakers.

[7]In cases where we were interested in the impact of multiple mixed effects and/or inter-individual variation, linear mixed models were applied. In other cases, where

figures for the different groups through relative (rather than absolute) numbers for non-canonical cases as a proportion of all relevant cases.

Morphology and Syntax

In the domain of morphology and syntax, examples come from the formation and use of participles in Russian, word order in German, and bare NPs in German.

Non-canonical Participles in Russian

Participles in Russian are challenging in their morphology and syntax, and they are acquired later by monolinguals (Cejtlin, 2009; Tribushinina et al., 2013), which makes them an ideal domain to look for non-canonical forms. Results of our corpus study show that morphologically non-canonical participles can be found across all speaker groups, including monolinguals:

(1) v′′ezžaščaja belaja mašina beloe avto (-) a:
driving.in white car white car
pritormozil
slowed.down

[DE**bi**39MR_fsR]

(canonical: v′′ezžajuščaja)

"A white car is driving in, the white car (-) a: slowed down."

(2) voditeli avtomobilej, kak pervogo, tak i sledovšego
drivers of.cars like first so also following
za nim, rezko pritormozili
after him abruptly slowed.down

[RU**mo**06MR_fwR]

(canonical: sledovavšego)

"The drivers of the cars, both the first one and the one that followed it, abruptly slowed down"

A qualitative analysis reveals interesting dynamics in the morphological formation of participles. With suffixes, there is a widespread truncation of material, which can be seen both in (1), produced by a HS, and (2), produced by a monolingual. In (1), the expected suffix *-juščˇ-* for the formation of active present participles of open stems ending with a *j*-addition (Bogdanov et al., 2009) is truncated to *-ščˇ-*. In (2), the base for the formation of the active past participle with the suffix *-vš-*, consisting of the stem and a thematic suffix *sled-ova-*"follow," is truncated to *sled-o-*. Such a pattern can be interpreted as a reduction of morphological complexity pointing at a tendency for stem unification across paradigms (Gagarina, 2002:160).

In order to check whether the frequencies of participles differ significantly across groups, we ran a one-tailed unpaired Wilcoxon rank sum test. Results show that HSs produced more non-canonical forms than monolinguals (M = 0.96, SE = 0.48): W = 36,480, p = 0.023 for HSs in the United States (M = 4.64, SE = 1.29), and W = 31,997, p = 0.032 for HSs in Germany (M = 4.11, SE = 1.28). Overall, HSs produced fewer participles, both canonical and non-canonical ones, than monolinguals (M = 0.69, SE = 0.09): W = 28,455, $p < 0.001$ for HSs in the United States (M = 0.15, SE = 0.04), and W = 26,118, $p < 0.001$ for HSs in Germany (M = 0.22, SE = 0.04), see **Figure 1** for

we simply wanted to look at frequency differences across groups, we used the non-parametric Wilcoxon rank sum test for non-normally distributed data and parametric ANOVAs for normally distributed data.

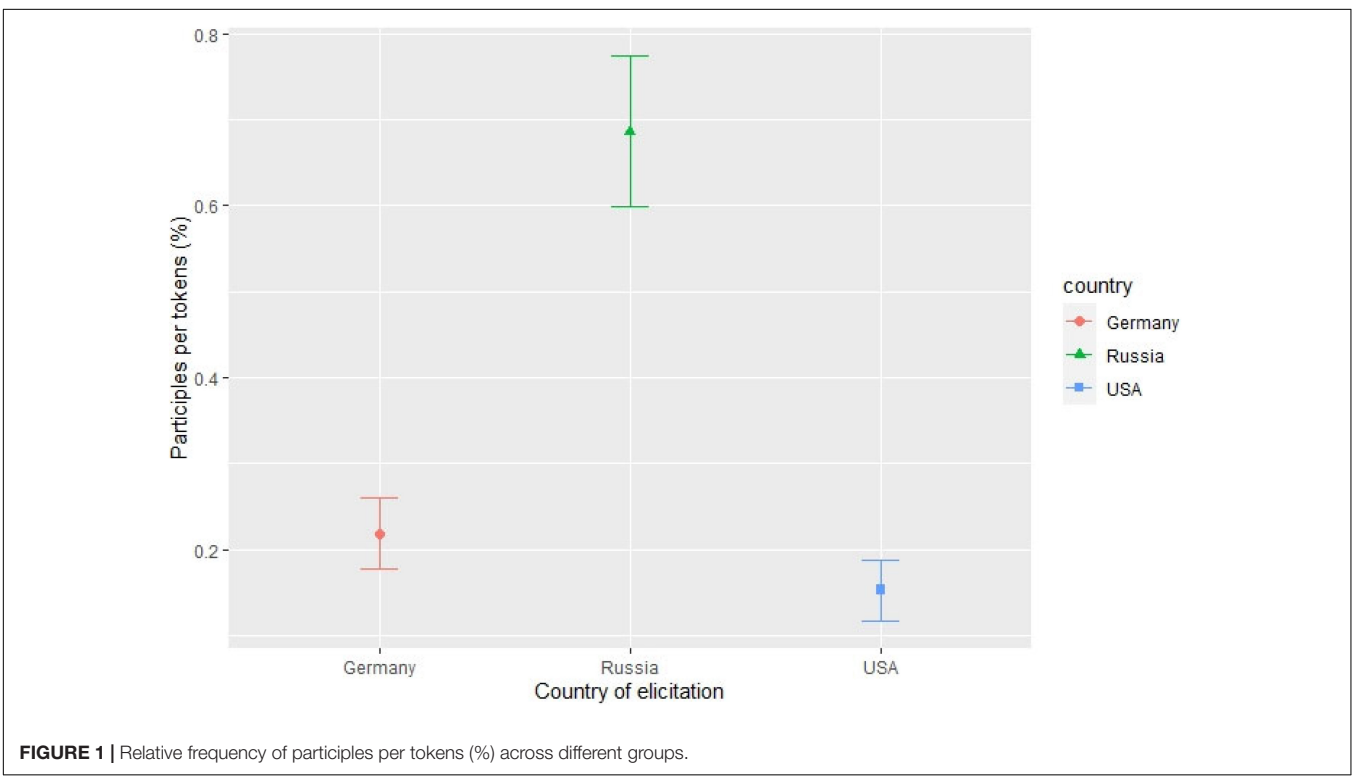

FIGURE 1 | Relative frequency of participles per tokens (%) across different groups.

TABLE 2 | Frequency of tokens and participles across different groups.

Country	Tokens: overall	Tokens: canonical participles (% of all tokens)	Tokens: non-canonical participles (% of all tokens)
DE	32,882	57 (0.17%)	14 (0.04%)
US	29,214	40 (0.13%)	14 (0.05%)
RU	25,930	151 (0.58%)	6 (0.02%)

the relative frequencies and **Table 2** for absolute frequencies of tokens and participles across the groups. Both findings might be explained by the abovementioned status of participles, and by the fact that they are generally rare in HSs' oral input, since they are associated with formal registers (Zemskaja, 1973; Golub, 2001).

Non-canonical participle formation is also well documented for monolingual child acquisition (Cejtlin, 2009). Interestingly, such forms often follow the same patterns as those we find in our data:

(3) Pečen'e uže s″eto. (Cejtlin, 2009:169)
cookies already eaten
(canonical: s″edeno)

"The cookies are already eaten"

Similar to (1) and (2), example (3) is a case of morphological truncation of a stem *s″ed-* "eat"- to *s″e-*, which then forms the base for participle formation with the suffix *-t-* rather than with canonical *-en*(*n*)*-*. The fact that we observed such patterns also in monolingual and bilingual (HS) adult speakers, suggests ongoing internal dynamics in this linguistic domain in Russian. That HSs use such a pattern with a higher frequency hence means that they can shed a spotlight on ongoing tendencies in native grammars.

Non-canonical Word Order in German

For German, we report relevant findings from two domains: word order and bare NPs. German has traditionally been described as an SOV language with verb-second (V2) word order requiring the finite verb in main declaratives to appear in second position, after exactly one constituent in the domain in front of it, the "forefield." This position of the finite verb constitutes one of two "sentence brackets" characteristic for the lay-out of German sentences. The other position is located at the right clausal periphery. In main declaratives, it contains non-finite verbs and separable verb particles. Together, the left and the right bracket delimit the "middle field," the canonical domain for complements and adjuncts. Typically, embedded clauses are extraposed, i.e., occur beyond the right sentential bracket in the "post-field."

The V2 requirement is usually regarded as a prime example of a rigid constraint in the grammar of German native speakers. Therefore, deviations from V2 in the maj-German of Turkish HSs, where an adverbial occurs in front of the subject at the left periphery, were taken to fall outside native German. For Auer (2013:37f), for instance, they indicated the reorganization of German V2 to SVO, which would "intervene deeply in the structures of autochthonous German in its standard and non-standard forms" and, together with other non-canonical patterns, such as bare NPs, "would have the potential to constitute a new variety that would differ substantially from autochthonous German" (German originals, our translation).

This view has been challenged by accounts integrating this non-canonical pattern into the syntactic lay-out of German sentences (Wiese, 2013; te Velde, 2017; Walkden, 2017; Wiese and Müller, 2018). Findings suggest a systematic verb-third (V3) option that, unlike SVO, preserves the characteristic German sentence brackets. From the point of view of information structure, V3 has the advantage over V2 of allowing both a framesetter or discourse linker [e.g., *dann* "then," see (4) and (5) below] and a topic in the left periphery (Wiese, 2012, 2013; Walkden, 2017; Wiese et al., 2017; Bunk, 2020).

While V3 in German has mostly been associated with language-contact situations (e.g., Walkden, 2017), we have shown that it is also available in monolingual speakers (Wiese and Rehbein, 2015; Bunk, 2020). The present study confirms this for maj-German across populations: in the RUEG corpus, we find V3 not only in bilingual speakers with h-Greek, h-Russian, and h-Turkish, but also in monolingual speakers, cf. (4) and (5).

(4) und dann er lässt sein ball einfach fallen
and then he lets his ball simply fall
[DE**bi**51MT_isD]

"And then, he just lets his ball fall down."

(5) dann die polizei is auch (-) richtig schnell gekommen
then the police has also really quickly come
[DE**mo**68FD_isD]

"And then, the police arrived really quickly."

Findings point to the same V3 options in monolinguals and bilinguals, with an adverbial and a subject preceding the finite verb. As in previous studies, V3 is infrequent, though, with only 48 cases in the bilingual and 11 cases in the monolingual group, with the bilingual group in the lead quantitatively (cf. also Wiese and Rehbein, 2015). In order to compare the difference, we computed normalized frequencies per 100 CUs (communicative units). In the RUEG corpus, CUs were used as a means to segment utterances and were defined as an "independent clause with its modifiers" (following Loban, 1976:9). Accordingly, normalizing for 100 CUs gives us the numbers for V3 as a percentage of all independent clauses, and hence a good approximation for the proportion of non-canonical cases, since V3 is a pattern located at the clausal level. The quantitative difference between speaker groups we observed for absolute numbers is confirmed for such normalized frequencies: we find 0.96 V3 occurrences per 100 CUs in the bilingual group, compared to 0.41 occurrences per 100 CUs in the monolingual group.

Interestingly, the higher frequency is primarily due to speakers with h-Turkish: data from this group makes up 38 of the 48 V3 findings, or 1.88 occurrences per 100 CUs, compared to only 0.42 and 0.27 occurrences per 100 CUs (five cases each) coming from

the h-Greek and h-Russian group, respectively. These findings speak against contact-linguistic transfer, since Turkish, an SOV language, would support the basic SOV word order of German, while Russian and Greek both have a tendency to SVO, which shares a strong surface similarity with V3 as soon as additional left-peripheral constituents such as adverbials are involved. If cross-linguistic transfer were a relevant trigger for V3, we would expect bilingual speakers with h-Greek and h-Russian to be in the lead, rather than those with h-Turkish. This further supports a view of German grammar as the locus of this phenomenon.

For h-German speakers in the United States, we observe an increase in the non-canonical V3 pattern, similar to what we found for non-canonical participle formation in h-Russian: h-German speakers produce 3.47 V3 main clauses per 100 CUs (55 cases in total), with framesetters/linkers like *dann*, *nun* in addition to those probably adopted from English, such as *so* ([zo]), cf. (6) and (7) (with the latter exhibiting all three linkers).

(6) dann um die ecke sind zwei autos gefahren
then around the corner are two cars driven
[USbi58FD_iwD]

"then two cars drove around the corner"

(7) und dann der auto hinter ihm war nicht bereit so
and then the car behind him was not ready so
der hat ihn (-) dreingeschubst
he has him there-in-pushed
und nun es is alles überall hingerollt
and now it is all everywhere rolled
[USbi62FD_isD]

"and then the car behind it wasn't ready so it pushed into it and now everything rolled everywhere"

In this case, the influence of English (X)SVO might further support V3 production, given that English is the majority and main contact language for h-German here. While we cannot rule out that some patterns are enhanced by parallels in English, this only holds partially, though, as shown in (6) and (7), where the clausal brackets remain canonical ("sind . . . gefahren;" "hat . . . dreingeschubst"), in contrast to English SVO. Qualitatively, the h-German data matches, to a large extent, what we find in monolingual and bilingual speakers of maj-German. The difference appears to be quantitative and due to an increased range of constituents involved in V3 clauses, such as non-subjects in the forefield [e.g., a PP as in (6)] and – trivially – borrowed linkers [such as *so* in (7)]. All other cases are attested in our monolingual data as well, even though with a lower frequency.

Had we only investigated h-German in contact with maj-English, we might have claimed that V3 is due to cross-linguistic transfer. Instead, we can now conclude these patterns are also available within the monolingual German repertoire but may be selectively strengthened in HSs by language contact.

Non-canonical Bare NPs in German

For the investigation of non-canonical bare NPs in German, we used additional data from the DNam corpus of German in Namibia (Wiese et al., 2017; Zimmer et al., 2020)[8], in order to compare two groups of h-German speakers. Namibian German represents a rare case of h-German that is still grounded in a vital speech community that systematically uses German not only in informal, but also in formal settings. We focus on the LangSit subcorpus which contains register-differentiated data similar to that in the RUEG corpus (103 speakers; 51,509 tokens). It covers informal-spoken and formal-spoken productions (formal: 23,606 tokens, informal: 25,629 tokens), elicited with visual stimuli in the form of a photo story about a car accident.

Our findings indicate that monolinguals as well as bilinguals produced non-canonical bare NPs, i.e., those that we would not expect in standard German. Figures are overall low: non-canonical cases make up 0.34% of all NPs across maj-German data by both monolingual and bilingual speakers in Germany (22 and 50 occurrences, respectively), 0.98% (22 cases) in h-German in the United States, and 1.06% (44 occurrences) in Namibia.

Interestingly, the non-canonical cases differ qualitatively between the h-German group in the United States and the others. Non-canonical NPs in Germany (maj-German by mono- and bilingual speakers) and Namibia (h-German) can be subsumed under current trends of article decline in German triggered by hyperdetermination, as described by Leiss (2010): (a) generally in generic and unique reference and light verb constructions and in local and directional contexts, and (b) a decline of the definite article in initial, thematic position, and of the indefinite article in rhematic position, because these are already inherently definite or indefinite, respectively. (8) illustrates this for a non-canonical bare NP with generic reference (and in rhematic position) from Namibian German, and (9) for one in rhematic position, produced by a monolingual speaker in Germany:

(8) dann hab ich auch krankenwagen angerufen
then have I also ambulance called
[D**NAM**_S_00066]

"Then, I also called (an) ambulance."

(9) ich hab eben verkehrsunfall beobachtet
I have just traffic.accident observed
[DE**mo**30FD_isD]

"I just observed (an) accident."

The h-German data in the United States differs from this in that we find a distinctive pattern that accounts for almost half of the cases (11 occurrences) and does not occur in the other data. In this pattern, non-canonical bare NPs form the second element in a coordination, cf. (10):

(10) Die blaue auto hat gehaltet wiel die ball und
the blue car has stopped because the ball and
hund war vorne
dog were in.front
[**US**bi08MD_fwD]

"The blue car stopped because there was the ball and (the) dog in front of it."

[8]http://agd.ids-mannheim.de/DNAM_extern.shtml

Unlike known patterns of determiner sharing (McCawley, 1993; Ackema and Szendrõi, 2002), this is not restricted to coordination at the VP level, suggesting that an existing pattern can be further extended in heritage language contexts. As the contrast to Namibian German shows, this variation is not related to heritage German *per se*, but might differ across speech communities. Hence, heritage languages can participate in ongoing tendencies as well as further extend native-grammar options.

Intonation

For the domain of intonation, our data provides an example from Russian yes-no questions (YNQs). In the literature, Russian YNQs are reported to be realized with a bitonal rising nuclear pitch accent on the verb (L* + H or L + H*; Rathcke, 2006b; Meyer and Mleinek, 2006) followed by a low final boundary tone (FBT; L%) (Igarashi, 2006; Rathcke, 2009), except if the nuclear pitch accent falls on the final syllable, in which case a high FBT (H%) is realized (Makarova, 2003; Rathcke, 2006a). The FBT can thus be considered truncated if no material follows the pitch accent (Rathcke, 2009, 2013). The intonation patterns of YNQ differ in Russian and English and are hence interesting to investigate in bilingual speakers.

In order to study the prosodic realization of YNQs of mono- and h-Russian speakers, we elicited experimental data in addition to the corpus data during data collection. This consisted of 10 read-aloud YNQs about details of the car accident. We recorded 20 speakers per group, i.e., (1) bilingual speakers of h-Russian in the United States, (2) bilingual speakers of h-Russian in Germany, and (3) monolingual speakers in Russia [see Zuban et al. (2020) and in prep. for details].

The elicited YNQs differed in the number of syllables following the nuclear pitch accent (or an additional pitch accent on the object for SVO questions). Each YNQ was annotated for the location of the nuclear pitch accent and FBT, following a combined phonetic and auditory approach: presence of a pitch accent was detected auditorily, and the FBT was examined with respect to local F0 trajectories and changes using Praat. Labeling followed the autosegmental-metrical framework (Makarova, 2003; Igarashi, 2006; Rathcke, 2006b).

Results of the descriptive analysis showed that h-Russian speakers in the United States predominantly produced L% (82% of all cases) while h-Russian speakers in Germany and monolinguals produced L% less frequently (58% for both groups).In order to check for a possible impact of multiple fixed effects on the distribution of high and low boundary tones, we ran a binomial generalized linear mixed-effects model with FBT as the dependent variable, and the three speaker groups, number of syllables following the last pitch accent to the FBT (0–5), transitivity, nuclear contour as independent variables, and with speaker and item as random effects (**Figure 2**; see "**Supplementary Appendix 1**" for full model specifications and summaries). It was found (among other things) that all speaker groups produced an H% when the last pitch accent fell on the final syllable, in line with the literature on Russian. Along the same lines, according to what was reported on truncation of FBTs in Standard Russian, h-Russian speakers in the United States chose the L% FBT as soon as there was at least one syllable following the last pitch accent. However, h-Russian speakers in Germany and mono-Russian speakers preferred the L% only when there were more syllables following the last pitch accent.

Hence it is only h-Russian speakers in the United States who behave according to what was described in the literature, while mono- and h-Russian speakers in Germany do not. Had we investigated only h-Russian in Germany and found the significantly increased use of H% (i.e., the absence of categorical truncation), we might have been led to think

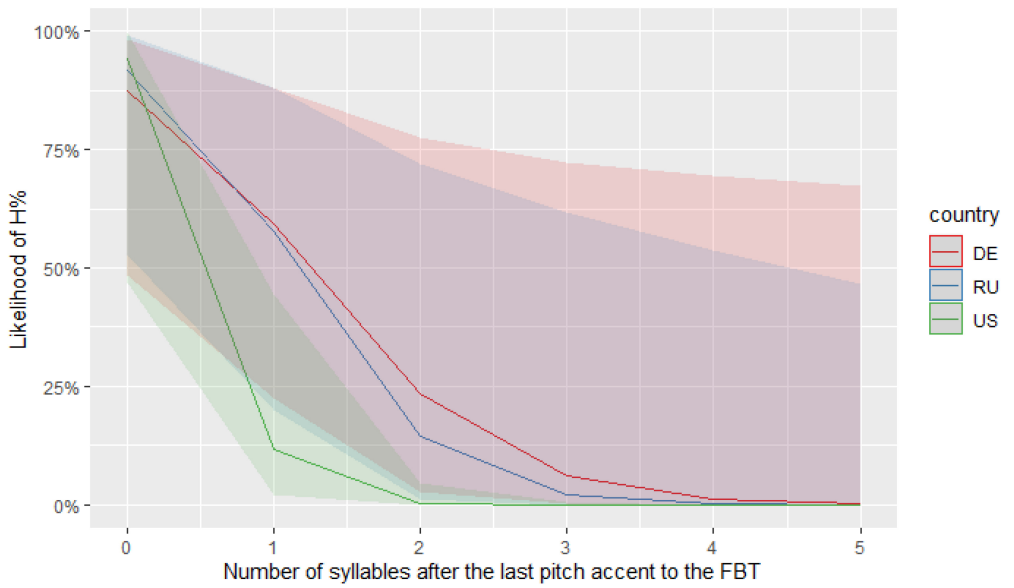

FIGURE 2 | Estimated likelihood of H% being chosen over L% based on the productions of the three speaker groups.

that this pattern is a specific feature of HSs' intonation grammar, possibly due to influence from maj-German, which has H% in YNQs (Grice et al., 2005). However, comparable realizations by mono-Russian speakers show that what we see here is a more general pattern: if we approach both HSs and monolinguals as native speakers and analyze language use across speaker groups, we may also find non-canonical patterns in monolinguals that might otherwise have been attributed to bilingualism.

Pragmatics

In the domain of pragmatics, examples come from data on the position of new referents in Turkish and on referent introduction in English.

Non-canonical Placement of New Referents in Turkish

As briefly mentioned in "Morphology and Syntax," the basic and pragmatically neutral word order in Turkish is SOV. However, for specific pragmatic purposes, elements can also be placed in the post-verbal position. In particular, this is possible for marking backgrounded information and afterthoughts (Erguvanli, 1984; Schroeder, 1995; Kornfilt, 1997), for constituents such as NPs, adverbs, discourse markers, forms of address, and finite subordinate clauses. What is crucial is that placement of new information is believed to be impossible here: according to the literature, the information placed in the post-verbal position has to be discourse-predictable or recoverable from previous discourse (Erguvanli, 1984:56).

Approaching HSs as native speakers of their languages, we investigated these information-structural restrictions for h-Turkish and mono-Turkish alike. We proceeded by selecting the 21 most frequently used nominal referents that played a role in the elicited narratives. We annotated every occurrence of these, adapting Riester and Baumann's (2017) referent annotation scheme, and identified information status through three categories (cf. Schroeder et al., to appear): (1) "new" (first mention), (2) "given" (referent that was introduced before) (3) "bridging" [referent that has not explicitly been introduced but belongs to the "pragmatic set" in the sense of Hawkins (1984) of a given referent (anchor)]. Afterthoughts, repairs and finite subordinate clauses were excluded from analysis.

Results of the analysis show that even though most of the referents placed post-verbally were indeed "given" and "bridging," there is a substantial number of new referents used in the post-verbal position. What is interesting in the context of the present manuscript, is that new referents in the post-verbal position occurred not only in the data of HSs [as seen in (11), but also in monolinguals in Turkey (12)]. In HSs in Germany, the occurrences of new referents in post-verbal position constitute 32.43% of the overall number of referents in the post-verbal position (24 new referents out of 74 referents in the post-verbal position), in monolinguals in Turkey the occurrences of new referents constitute 22.45% (11 new ones out of 49 referents), and in HSs in the United States, the new referents in the post-verbal position make up 21.43% (24 occurrences out of 112). This outcome contradicts what the literature says about canonical Turkish, namely that new referents are not possible in the post-verbal position.

(11) bi: çift vardı çocuk arabasıyla
one couple was child car.with
[DE**bi**18MT_fsT]

"There was a couple with a stroller."

(12) az kalsın bi araba çarpıyodu çocuklu bi kadına
almost one car hit child.with one woman
[TU**mo**26MT_isT]

"A car almost hit a woman with a child."

Most of the new referents in post-verbal position stand in a close semantic relationship to the subject of the clause. This relationship is wider than the "bridging" relation, and it is often indicated by means of the possessive suffix *-(s)I* on the post-verbal constituent [like in (11)], or by free adjuncts that carry adverbial case. Less often, a new referent in post-verbal position is a lexical subject or object [as in (12)].

However, it would be misleading to call those new referents that are in a close semantic relationship with the subject "backgrounded information," since in about half of the cases in our data, the referent introduced in the post-verbal position is mentioned again in the subsequent discourse, and is treated as "given" when mentioned again. We propose to call such referents "secondary" new referents, in the sense that they are secondary (and related) to another new referent with a higher relevance to the discourse at that point.

Thus, we conclude that it is indeed possible to place new information in post-verbal position in Turkish, and this is not a feature that is typical only for HSs, since the pattern is also found in the monolingual data from Turkey. As we will discuss in more detail in section "Register Leveling in Heritage Languages," in monolinguals this pattern seems to be associated with informal registers. Hence, if we compare like with like and systematically include data from such registers from monolinguals as well, we can avoid misattributing some non-canonical patterns to bilingualism that form a more general part of native grammars.

Non-canonical Referent Introduction in English

Unlike Turkish, English marks newness and givenness of referents through indefinite and definite articles (Hickmann and Hendriks, 1999). The indefinite article *a* presupposes that the referent of the NP is new and the addressee is not familiar with it. The definite article *the* implies that the addressee can uniquely identify the given referent of the NP based on previous discourse, shared physical environment and/or general world knowledge (Payne and Huddleston, 2002:368–371).

Previous research has shown that bilingual speakers often differ in their production of articles from monolingual speakers of English. For example, child HSs of other languages have been reported to oversupply *the* in indefinite contexts and *a* in definite contexts in maj-English, regardless of their heritage language (Zdorenko and Paradis, 2008, 2012). Further, adult L2 English speakers with article-less L1s tend to overuse *a* and *the* in contexts with mismatching parameters of definiteness and specificity (i.e.,

subjective noteworthiness of the referent to the speaker) (Ionin, 2006; Ionin et al., 2008).

Most importantly, many studies on article use compare only two or more bilingual groups to each other (Lardiere et al., 2004; Zdorenko and Paradis, 2008, 2012; Ionin and Díez-Bedmar, 2021). If a monolingual comparison group is added, monolinguals usually supply articles in strict accordance with the expectations based on the literature (Hawkins et al., 2006; Ionin, 2006; Ionin et al., 2008; Sarko, 2009; Snape et al., 2013). Overall, these studies seem to suggest that variability in article production is a result of bilingualism.

We tested this assumption by examining article choice in new and given referents among bilingual and monolingual English speakers of maj-English – 214 HSs with various heritage languages and 64 English monolinguals.[9] Similarly to the investigation of new postverbal referents in Turkish just discussed, we selected 19 frequent referents such as *man*, *dog*, *car1*, and *car2*, and coded them for their information status as "new" (the first mention of an entity without any identifying information) or "given" (all the subsequent mentions) (Riester and Baumann, 2017). This yielded 4,961 new and 10,881 given referents.

We identified all new and given referents that were part of unexpected non-canonical structures, that is, "*the* + new" and "*a* + given" referents. All other structures in which referents appeared, including expected canonical structures ("*a* + new" and "*the* + given"), were marked as "other." Contrary to what would be expected from the literature, we found non-canonical patterns not only in bilinguals' productions [see (13) and (15)], but also in those of monolinguals [see (14) and (16)], and we found this for both the "*the* + new" pattern [(13) and (14)] and the "*a* + given" pattern [(15) and (16)]:

(13) Oh my god, I just saw a car get rear-ended, there was this really cute dog and he ran out in front of it because this guy dropped his ball and the dog chased it onto the street. And then they had to pick up the groceries [*new*]. [US**bi**50FD_isE]

(14) I'm calling about incident number F16, I was there at the time it happened. There was a blue car and a white car both coming down a path, they both made a right. As they made a right, a man had a ball, and it went into the street into the pathway. The dog [*new*] ran after the ball. [US**mo**07FE_fsE]

(15) There was this couple with a ball on one side and a lady packing groceries on another. The dude from the couple was holding a ball [*given*]. [US**bi**03MG_iwE]

(16) Then a dog starts barking and he runs in the street. So this guy is driving down the street and he sees a dog [*given*]. [US**mo**10ME_isE]

Hence, we did not find any qualitative differences between monolinguals and bilinguals in this domain. In order to check for a possible impact of bilingualism on quantitative distributions while taking into account inter-individual speaker variation, we ran two binomial generalized linear mixed effects models, one for new and one for given referents. The dependent variable was Determiner ("*the* + new"/"*a* + given" vs. "other") and the independent variables were Bilingualism (bilingual vs. monolingual), Setting (formal vs. informal) and Mode (spoken vs. written) (see "**Supplementary Appendix 2**" for full model specifications and summaries).

The results indicate neither a main effect of bilingualism, nor its interactions with other variables, meaning that we have no evidence of differences between HSs and monolinguals. Both groups produced similar numbers of non-canonical "*the* + new" referents, ranging from 6.6% of all given referents by heritage speakers and 5.6% by monolinguals in the formal written situation, to 9.1% by heritage speakers and 11.1% by monolinguals in the informal spoken situation. The two groups did not differ in the production of "*a* + given" referents either: for this pattern, the percentages of non-canonical referents were much smaller and ranged from 0.47% of all new referents by heritage speakers and a complete absence by monolinguals in the informal written situation, to 0.75% in the informal spoken situation by heritage speakers and 1.5% by monolinguals. Overall, our data shows that a pattern that has mostly been attributed to bilingualism actually manifests itself in the productions of English monolinguals in the same way as it does in the speech of bilinguals.

Variation: Not Just in Heritage Speakers, but Also in Monolinguals

In the literature, HSs and bilingual speakers in general are often presumed to exhibit higher degrees of variation than "regular" native speakers. Seton and Schmid (2016:341) even claim that "[t]he most striking characteristic that sets bilinguals apart from monolinguals is a larger amount of variability in performance." Observations from our data suggest a more complex picture and point to intricate dynamics between groups and individual effects and different degrees of dynamicity in linguistic subsystems. The evidence provided to illustrate this point here comes from our corpus data on maj-German, spoken by monolingual and bilingual speakers in Germany (see **Table 1** above).

Variation can be measured in different quantities, or in different qualities, such as a wider range of structures. For example, a bilingual speaker may use different structures from a monolingual, and those may be canonical or non-canonical (some canonical structures may be dispreferred by some speakers and/or in some registers). In our data, we find that all syntactic dependencies, such as different types of objects, particles, modifiers, etc., are used by mono- and bilinguals alike (see **Figure 3**). Hence, we do not find qualitative differences here: bilinguals (as a group) neither avoid certain dependencies used by monolinguals, nor do they exhibit a wider range of dependencies than monolinguals[10].

[9]See figures for maj-English in **Table 1** above. Eight HSs (three with h-German, two with h-Greek, and three with h-Turkish) were excluded from the analysis due to technical reasons.

[10]Even though **Figure 3** contains some fine-grained prepositional subcategories (PNA and PND marking accusative/dative complements to prepositions) that go beyond the underlying dependency grammar (Foth, 2006), and dependency parses have been manually corrected, it is possible that more diversity on this level

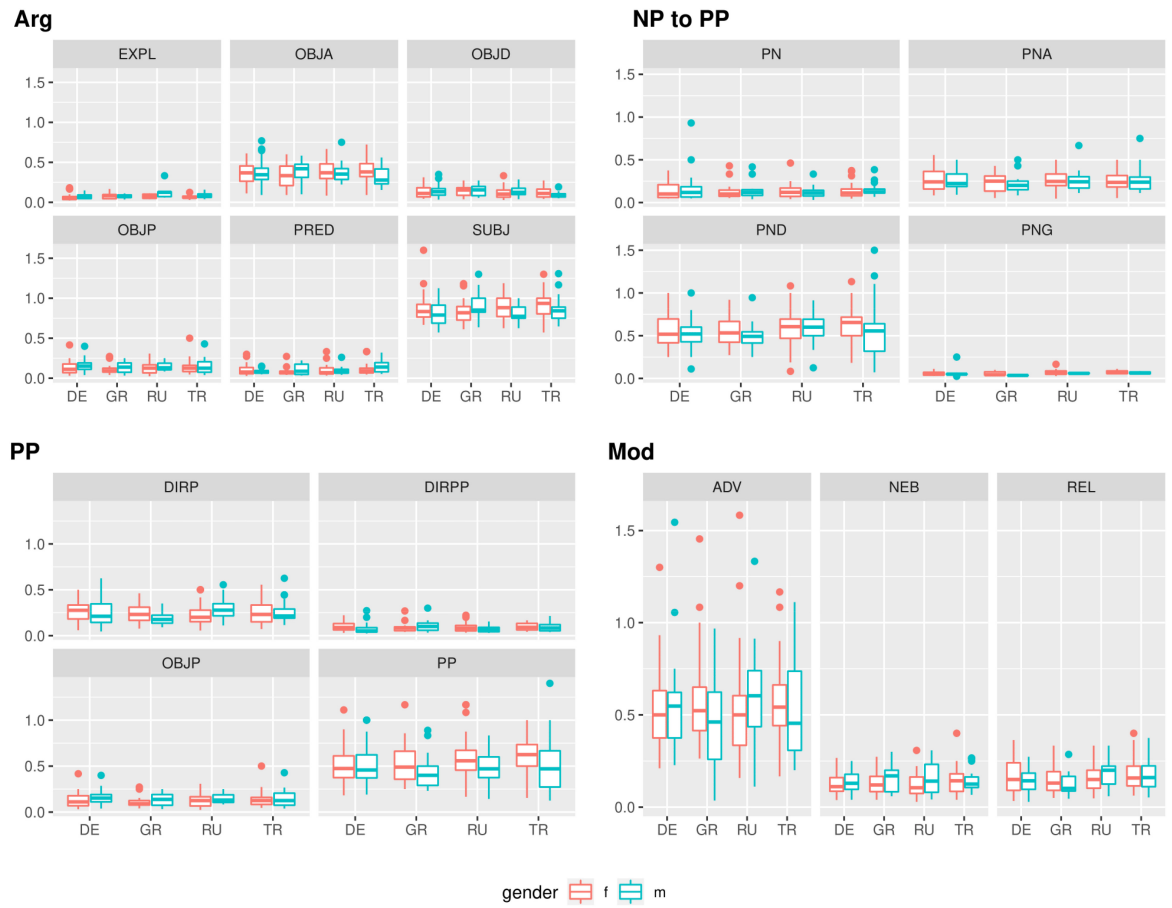

FIGURE 3 | Dependencies in maj-German (formal-written) RUEG texts, normalized by lexical verbs.

Are there quantitative differences in variability, i.e., differences in variance, then? The picture here is more mixed. First of all, both the mono- and the bilingual speakers show large degrees of quantitative variance for some, but not all structures. This is in part an effect of the general frequency of a structure: if a dependency type is overall rare, its frequency of occurrence will exhibit floor effects. However, it also appears that some structures, whether rare or more frequent, are subject to more free choice in the frequency of their realization. This can be due to simple causes, such as the possibility to name all vs. just some of the agents or circumstances in the respective narratives, or recurrence and repetition of previously named entities vs. ellipsis. At the same time, a more or less frequent realization of elements may also be indicative of differences in communicative style (explicitness, emphasis) or formality (assumption of interlocutor expectations; see also Ahern et al., 2019: 487, 488). This can, for example, affect the frequency of modifying structures such as adverbs or relative clauses, which can be used to specify or comment on another dependency. Explicitness, referring here to the tendency to explicitly name more aspects of the environment or circumstances, may also play a role for different amounts of PP realizations.

However, the realization of PPs can also be due to interactions with the lexicon, for example in the case of prepositional objects required by certain verbs (abbreviated as OBJP – a full list of abbreviations is provided in "**Supplementary Appendix 3**"). Other PPs can be added attributively without lexical constraints. Differences in the frequency of realization of nominal complements to prepositions (PND for dative objects to prepositions, PNA for accusative objects complementing prepositions) can thus be due to lexical diversity (different verbs requiring different complements), but might also point toward case dynamics (identical verbs occurring with different complement cases). Differences in variance can thus represent very different types of phenomena.

In our data, we see variable degrees of variance between structures, but there is barely any evidence for a higher degree of variability in bilinguals. Only the h-Turkish speakers appear to show a higher degree of variance in the realization of (free) prepositional phrases (PP) and dative complements to prepositional phrases (PND), and the effect is rather small. Some bilingual groups appear to show a slightly higher variance in modifying structures (male h-Russian and male h-Greek

only becomes visible under consideration of more fine-grained lexicosyntactic phenomena such as constructions.

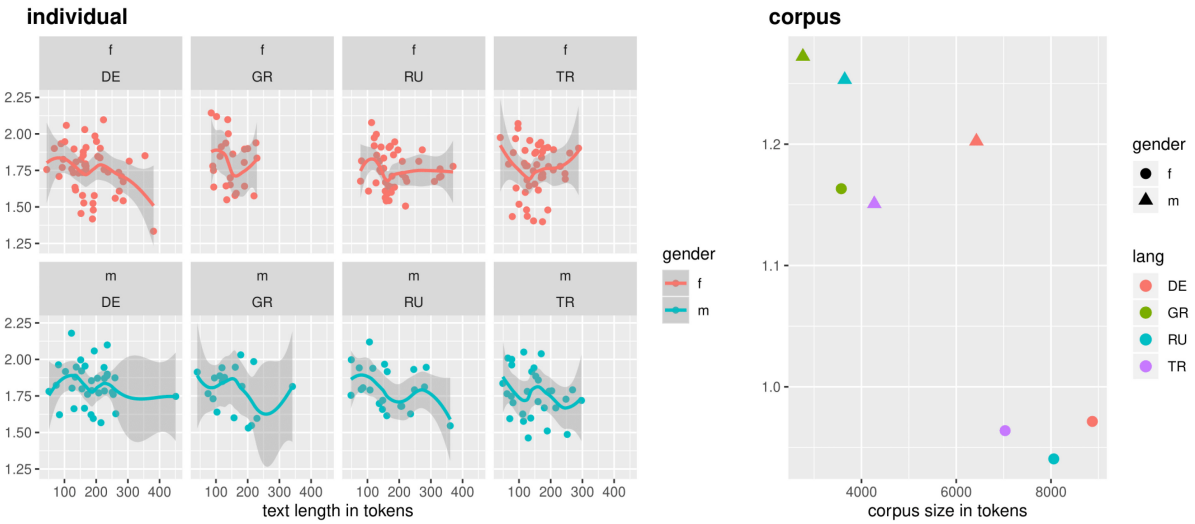

FIGURE 4 | Transformed TTR in maj-German (formal-written) RUEG texts.

for subordinate clauses, marked NEB, male h-Greek and male h-Turkish for adverbs, ADV). However, some chance results are to be expected due to the high number of between-group comparisons.

Overall, there is no clear trend toward higher variability in bilinguals: mono-German speakers are generally within the range of variance found for bilingual speakers, and even have a tendency to be on the upper end of it. In fact, differences between gender groups are generally higher than between speaker groups of different language background but the same gender.

This also appears to be the case for lexical richness, approximated here with a transformed type-token-ratio (TTR, see **Figure 4**)[11]. We find no strong differences in the distribution of individual speakers in each language group. However, the female mono-German and female h-Turkish bilinguals show the highest degrees of variance in TTR. This may be an artifact of different group sizes (23 female, 10 male) – especially since, in general, we find higher variance in the male groups. This becomes particularly obvious in the distribution by subcorpora: we find much higher TTRs for the male speaker groups compared to the female ones *across language groups*. Male and female speakers use roughly the same amount of lexemes per speaker, but those lexemes converge less between male than female speakers. This suggests divergent effects of idiomaticity or coselectional constraints[12] and might be attributed to communicative style or a higher degree of adjustment to assumptions of interlocutor expectations.

Overall, we find different degrees of variance in our data, but no particular effect for monolinguals vs. bilinguals. In fact, the monolingual speakers in the RUEG corpus exhibit a degree of variance that reaches, and sometimes surpasses, that of the bilingual speakers. We do not find evidence that variability in the realization of syntactic structures or in lexical richness could be used as a criterion for the definition of a "real" native speaker that would favor monolingual over bilingual speakers.

The Role of Registers

Approaching HSs as native speakers of both their languages, we targeted formal and informal registers in bilinguals and monolinguals alike. Our findings indicate an important role of registers for non-canonical patterns, and this can play out differently in heritage and majority language use. Below we discuss evidence from English and German as majority languages ("Noncanonical Phenomena in Informal or Spoken Registers of Majority Languages"), and German, Turkish, and Greek as heritage languages ("Register Leveling in Heritage Languages").

Non-canonical Phenomena in Informal or Spoken Registers of Majority Languages

In majority language use, non-canonical patterns are dominantly found in informal and/or spoken registers, and this holds across bilingual and monolingual speaker groups.

Association With Spoken Registers for Non-canonical Patterns in English

In section "Pragmatics" we discussed two non-canonical patterns involving new and given referents in maj-English, namely "*a* + given" and "*the* + new" referents, which appear in bilingual and monolingual English speakers alike. Our further results reveal the influence of spoken vs. written mode on non-canonical article choice. In the two linear mixed effects models reported in section "Pragmatics," we observed a main effect of mode across speaker groups and formal/informal settings: the spoken mode showed

[11]The TTR is transformed by the formula types/tokens × 4√tokens, or TTR × 4th root of text length (or corpus size, respectively) and serves only to adjust for text-length dependency. No claims are made to the general usefulness of approximating lexical richness through TTRs.

[12]See Shadrova (2020) for a discussion of measuring coselectional constraints in corpora.

more non-canonical structures than the written mode (for "*the* + new": $mean_{formalspoken}$ = 7.68%, $mean_{formalwritten}$ = 6.12%, $mean_{informalspoken}$ = 10.1%, $mean_{informalwritten}$ = 7.3%; for "*a* + given": $mean_{formalspoken}$ = 0.69%, $mean_{formalwritten}$ = 0.57%, $mean_{informalspoken}$ = 1.11%, $mean_{informalwritten}$ = 0.23%; see "**Supplementary Appendix 2**" for *p*-values and model summaries).

The higher variability in article choice in the spoken mode could be associated with a higher cognitive load of online (spoken) productions compared to offline (written) ones. Spontaneous spoken production often exerts performance pressure since it allows little time for planning and no possibility for changing what has been said (Pullum and Huddleston, 2002:12). This factor is important in L2 research: for instance, Ionin et al. (2021) argue for testing article knowledge of L2 speakers in comprehension rather than production in order to avoid the performance pressure and to evaluate speakers' implicit sensitivity to (in)definiteness. In our study, performance pressure in the spoken mode might have led the speakers to only consider their own perspective (familiarity with the referent) and, consequently, use the definite article, while ignoring the addressee's perspective (unfamiliarity with a new referent), which would require the indefinite article.

In addition, we might be witnessing a new development in English: possibly, the definiteness distinction is becoming less strict in spoken spontaneous productions. So far, it is unclear if this is a systematic pattern of English internal dynamics (since over 90% of *the* uses are still canonical), and it needs to be confirmed in future research.

Association With Informal Registers for Non-canonical Patterns in German

In maj-German, non-canonical V3 (as described in "Morphology and Syntax" above) occurs mainly in the informal productions in our corpus data: of 59 V3 cases in maj-German altogether, 53, that is, roughly 90%, are from informal communicative situations, and this pattern is also evident for normalized frequencies, with 1.24 of 1.37 occurrences per 100 CUs, that is, roughly 91%, from informal communicative situations.

Within the informal settings, we find V3 across spoken (17) and written (18) modes. This suggests that unlike non-canonical article choice in maj-English, V3 in maj-German is primarily associated with informality, independently of mode.

(17) und dann er lässt sein ball einfach fallen
and then he lets his ball simply fall
[DEbi51MT_**is**D]

"And then, he just lets his ball fall down."

(18) danach er lässt den ball aus versehen fallen
afterward he lets the ball from accident fall
[DEbi58MT_**iw**D]

"Afterward, he accidentally lets the ball fall down."

Note, though, that we do find a few occurrences of V3 in formal data, and this also includes one production (formal-written) from a monolingual speaker:

(19) am helligten tage ein paar mit kinderwagen
at.the bright.light day a couple with stroller
sowie einem fußball ging auf dem fußweg
as.well.as a foot.ball went on the foot.path
[DEmo31MD_**fw**D]

"In bright daylight, a couple with a stroller went along the sidewalk."

This, again, integrates bilingual and monolingual speakers alike into the native speaker group: we find the non-canonical V3 pattern across groups, in both groups primarily associated with informal registers, with some exceptions which are also evident in both groups.

For non-canonical bare NPs in maj-German, we similarly found a dominance in informal data, and again this held for bilinguals as well as monolinguals. We also found this pattern in

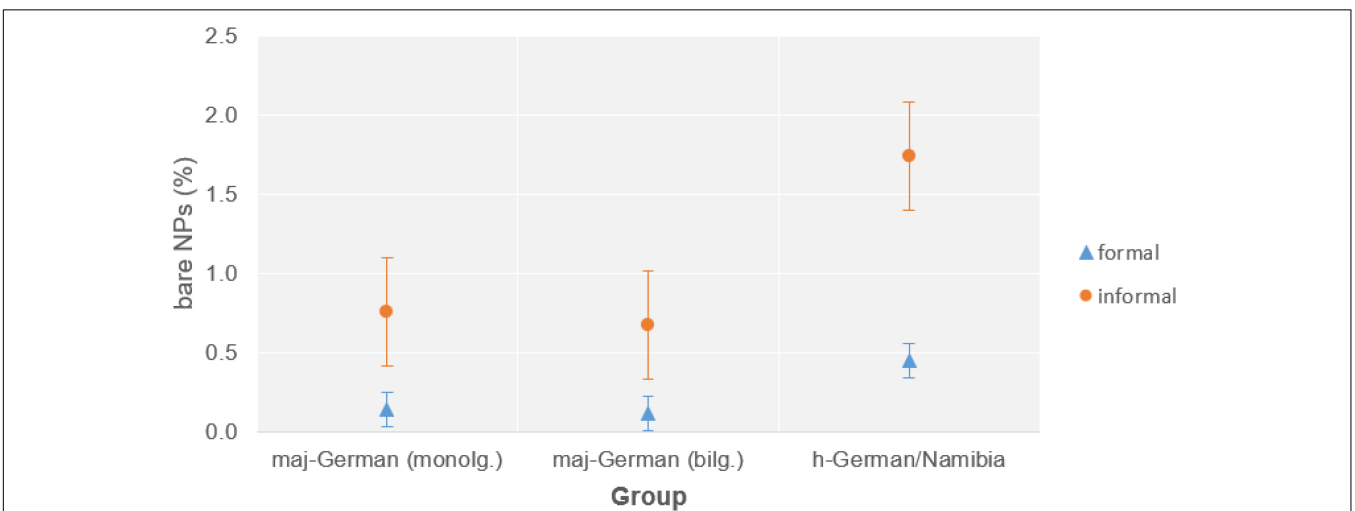

FIGURE 5 | Relative frequencies of non-canonical bare NPs (percentage of all NPs) in German in informal vs. formal settings by three different speaker groups.

h-German in Namibia, where it was even more pronounced, cf. **Figure 5**,[13] suggesting that such register differentiation can also hold in heritage grammars (but see "Register Leveling in Heritage Languages" below for h-German in the United States).

Register Leveling in Heritage Languages

Some patterns characteristic for informal registers in monolinguals can be generalized to formal registers in bilinguals' heritage language use, leading to register leveling.

Non-canonical Bare NPs and Word Order in h-German in the United States

One example for this is h-German in the United States In the United States, speakers do not distinguish between formal and informal settings in their non-canonical word order and NP patterns, unlike (monolingual and bilingual) speakers in Germany.

For h-German V3, we find 1.89 occurrences per 100 CUs in formal settings and 1.57 in informal settings (30 and 25 occurrences, respectively). Mode, however, plays a role here, as 71% of all V3 clauses appear in the spoken mode, pointing to a more general phenomenon that we also saw in non-canonical article choice in maj-English.

H-German speakers in the United States also show a similar distribution of non-canonical bare NPs in formal and informal situations, with 0.15 and 0.17% of all NPs (13 and 10 occurrences), respectively. As the data discussed in "Noncanonical Phenomena in Informal or Spoken Registers of Majority Languages" shows, this is in contrast not only to monolingual and bilingual speakers of maj-German in Germany, but also to h-German speakers in Namibia, where non-canonical NPs were primarily associated with informal registers. This can be related to the different communicative domains for German: in Germany and Namibia, German is used in informal as well as formal settings, including formal schooling, whereas in the United States, it is mostly restricted to informal contexts.

Further evidence for register leveling comes from h-Turkish in Germany. In section "Pragmatics," we found non-canonical placement of new referents in post-verbal position in both HSs and monolinguals. Our analysis shows that in monolingual data, all the examples of new referents in the post-verbal position occur in informal (spoken and written) settings. HSs, on the other hand, tend to place new referents in the post-verbal position across all communicative situations. Again, this can be related to communicative domains for the heritage language: similarly as for h-German in the United States, h-Turkish speakers in Germany are mostly exposed to informal language at home and with their peers and have less access to formal registers. A phenomenon typical for informal settings in mono-Turkish then spreads into formal settings in h-Turkish.

These findings, then, emphasize that non-canonical phenomena should not always be attributed to heritage languages or bilingualism in general, but they can also be a feature of informal settings in monolingual varieties. Thus, what might been seen as a consequence of language contact or attrition in a superficial account turns out to be present also in monolingual language use when different communicative situations are taken into account.

Non-canonical Restrictive Relative Clauses in maj- and h-Greek

A further area where we noted register leveling is the distribution of restrictive relative clauses (RCs) in Greek. RCs in Greek function as modifiers of nouns, as in other languages, and always appear in post-nominal position (Chatsiou, 2010). Greek RCs[14] come in two types: they are either introduced by the pronoun *o opios*, literally "the who," or by the complementizer *pu* "that." While the pronoun is inflected for gender, case and number and agrees with the nominal head that it modifies in gender and number, *pu* bears no inflection. In the literature on Greek, we find the claim that *pu* appears mostly in colloquial speech, while the pronoun is preferred in formal registers (Holton et al., 1997: 212).

We investigated restrictive RCs for the three groups represented in our corpus: monolingual speakers of maj-Greek in Greece and bilingual speakers of h-Greek in Germany and in the United States. As **Table 3** shows, *pu* RCs are more frequent than *o opios* RCs across the three speaker groups in the different countries, and this holds in both informal and formal settings.

Zooming in on the question of register variation, it appears that monolingual speakers actually prefer *pu* RCs even in formal settings, something that contradicts the claims in grammars of Standard Modern Greek. This points to a process of leveling even in the monolingual group. A similar pattern is exhibited by the h-groups, who also favor *pu* RCs over *o opios* RCs across settings, and a one-way ANOVA test[15] revealed no significant differences between the three groups concerning the production of *pu* RCs in the different communicative settings.

For *o opio*s RCs, we find differences in the distribution across registers between groups, as determined by one-way ANOVA tests {formal: $[F(2,171) = 15.99, p < 0.0001]$, informal: $[F(2,171) = 8.877, p < 0.0001]$}. Our data indicates that the *o opios* strategy is preferred in the formal register compared to the informal one (although at a lower rate than *pu* RCs). In this case, there is a statistically significant difference between

[13]Note, though, that the corpus data for h-German in Namibia comes only from spoken productions (see "Morphology and Syntax" above).

[14]Free RCs are excluded from this account.

[15]According to the central limit theorem, the sample distribution will be approximately normally distributed as the sample size is quite large ($N > 50$ items).

TABLE 3 | Quantitative data - distribution of restrictive RCs in different registers per group.

Communicative situation		h-Greek in the U.S.	h-Greek in Germany	Mono-Greek
Informal	*Pu*	165 (99.4%)	150 (82%)	183 (74.7%)
	o opios	1 (0.6%)	33 (18%)	62 (26.3%)
Formal	*Pu*	238 (97.9%)	190 (71.7%)	244 (64%)
	o opios	5 (2.1%)	75 (28.3%)	137 (36%)

groups, as determined by a one-way ANOVA [$F(2, 171) = 14.50$, $p < 0.0001$]. Specifically, Bonferroni-adjusted *post hoc* tests enabling pairwise comparisons revealed that the United States. group differs from the other two in the production of *o opios* RCs across the different communicative settings ($p < 0.05$). The preference for formal registers holds in particular for the monolingual group, and secondly for HSs in Germany. HSs in the United States rarely use *o opios* RCs regardless of setting [as reported also by Lithoksoou (2019), who investigated part of RUEG corpus].

Taken together, our results indicate register leveling for *pu* RCs across monolingual and bilingual speaker groups, and differences for *o opio*s RCs between different heritage speaker communities, with those in Germany aligning with monolingual speakers in Greece.

Such findings underline the relevance of considering register differentiations in bilinguals and monolinguals alike, while integrating HSs into the native speaker continuum.

DISCUSSION

In this manuscript, we argued for a perspective on bilingual heritage speakers as native speakers of both their languages, and demonstrated what can be gained from such an approach. We showed that recognizing heritage speakers as native speakers is justified on conceptual and theoretical grounds, and that this perspective is also a better fit for the empirical data. We presented results from a large-scale, cross-linguistic study that approached bilingual heritage speakers and monolingual speakers on equal terms, rather than using monolinguals as a yardstick for what counts as a competent "native speaker." In line with this approach, we targeted actual language use that covered broader repertoires than just formal language, and we did so for bilingual and monolingual speakers alike. To this end, we elicited linguistic productions representative of speakers' natural behavior in formal and informal, written and spoken communicative situations. This provided us with comparable data across registers, languages, contact-linguistic settings, and speaker groups, incorporated in an open-access corpus, the RUEG corpus.

Our findings support the integration of heritage speakers into the native-speaker continuum and show that they can shed light on language variation and change in native grammars. In a number of heritage languages, we find patterns in formal registers that do not appear in the respective majority settings for those languages. However, a closer look shows that these patterns are by no means completely absent from monolingual grammars, but can be associated with informal registers there. Patterns that might otherwise have been attributed to bilingualism could hence inform us on ongoing developments and variation in native grammars.

Our data points to a range of non-canonical patterns in monolinguals' productions that have so far been considered absent from native grammars. We discussed non-canonical patterns of participle formation and boundary tones in Russian; verb-third and bare NPs in German; new referents in post-verbal position in Turkish; and non-canonical choices of (in-)definite articles for given vs. new referents in English. Our study showed that these patterns are not restricted to heritage speakers, but they occur systematically in monolinguals as well.

Along similar lines, we found inter-speaker variability not restricted to bilinguals either. In our data for German, monolinguals displayed a degree of lexical and morphosyntactic variation that was sometimes higher than that of bilinguals, further challenging the model of the streamlined native speaker.

In majority language use, non-canonical patterns were dominant in spoken and/or informal registers, and this was true for monolinguals and bilinguals alike. In majority-English, for example, spoken registers featured more non-canonical article choices than written registers in both monolingual and bilingual speakers, and in majority-German, non-canonical word order and bare NPs were associated with informal registers.

In several languages, though, our data points to tendencies of register leveling in heritage contexts. This is presumably due to a lack of formal schooling: heritage speakers are mostly exposed to informal language at home and often have limited access to formal registers. Hence, phenomena typical for informal settings in majority language use can spread to formal settings in heritage languages. This points to different register distributions rather than to distinct grammatical patterns in heritage speakers. Accordingly, where the heritage language is also used in formal schooling, e.g., in the case of heritage German in Namibia, patterns were similar to those found for majority language use (i.e., German in Germany). As our Greek data showed, register leveling can also occur in monolingual speakers: we found that, contrary to claims in the literature, a non-inflected relativizer is preferred in formal as well as informal registers, and we found this for monolinguals and heritage speakers alike.

Some non-canonical patterns in informal settings were more frequent in bilingual speakers. Heritage speakers can hence put a unique spotlight on internal developments. In line with this, several of the non-canonical patterns we found point to extensions of existing, salient variants and/or ongoing language change. This underscores the importance of taking into account a broader range of communicative settings, not just for heritage speakers, but also for monolinguals.

For heritage languages, our findings also indicate that the size and cohesion of speech communities can play a role. In smaller, more widely distributed communities like our populations in the United States, we sometimes found more diverging patterns. This does not imply that those options are outside native grammars, but we might see more variation. Our data point to broader options of non-canonical word order and bare NPs in heritage German in the United States, while heritage German in Namibia patterned with majority German. For heritage Russian, we found a lower frequency of participles in the United States, while heritage Russian in Germany patterns with majority Russian.

Our findings also lead us to reconsider the state-of-the art on majority languages, given recurring evidence for non-canonical patterns in monolinguals that deviate from what has been

assumed in the literature so far. In the case of the intonational pattern analyzed in Russian, it was in fact the "deviant" pattern, that is, the one that differed from those described in the literature, that we found in monolinguals as well as in one of our heritage language communities, namely in Germany, but not in the other, that is, in heritage Russian in the United States.

Taken together, our findings support current calls to normalize multilingualism. Multilingualism can act as a motor of linguistic developments, and accordingly, multilingual communities can afford us a privileged view into ongoing tendencies of language variation and change. However, this does not make them an exotic, special case. Our findings put multilinguals' language solidly within native grammars, at levels of structure as well as language use. In order to make full use of the opportunity that multilingual speakers provide for linguistic theory, we need to take into account variation in native grammars, including informal registers, in bilinguals and monolinguals alike.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://doi.org/10.5281/zenodo.3236069.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by DGfS Ethics Committee (Deutsche Gesellschaft für Sprachwissenschaft/German Society for Linguistics). Written informed consent to participate in this study was provided by participants or, in case of minors, their legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

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

FUNDING

Research for this volume was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit "Emerging Grammars in Language Contact Situations" (AL 554/13-1, 15-1; AL 1886/2-1, 3-1, 4-1; BU 2690/1-1; GA1424/7-1, 10-1; LU 856/16-1; SCHR 1261/3-1, 4-1; SH 1685/1-1; SZ 263/4-1, 6-1; TR 238/5-1, 6-1, 7-1; WI 2155/10-1, 10-2, 11-1, 12-1, 13-1; and ZE 940/2-1, 4-1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717973/full#supplementary-material

REFERENCES

Aalberse, S., Backus, A., and Muysken, P. (2019). *Heritage Languages. A Language Contact Approach*. Amsterdam: Benjamins.

Ackema, P., and Szendrōi, K. (2002). Determiner sharing as an instance of dependent ellipsis. *J. Comp. Ger. Linguist.* 5, 3–34.

Ahern, A., Amenós-Pons, J., and Guijarro-Fuentes, P. (2019). Interpreting mood choice effects in L2 and L1 Spanish: empirical evidence and theoretical implications. *Appl. Linguist. Rev.* 10, 469–490. doi: 10.1515/applirev-2017-0097

Auer, P. (2013). "Ethnische marker zwischen varietät und stil," in *Das Deutsch der Migranten*, ed. A. Deppermann (Berlin: de Gruyter), 9–40.

Bayram, F., Kupisch, T., Pascual y Cabo, D., and Rothman, J. (2019). Terminology matters on theoretical grounds, too! Coherent grammars cannot be incomplete. *Stud. Second Lang. Acquis.* 41, 257–264. doi: 10.1017/S0272263119000287

Bogdanov, S. I., Evtjuxin, V. B., Knjazev, J. P., et al. (2009). *Morfologija Sovremennogo Russkogo Jazyka: Uèebnik Dlja Vyssix Uèebnyx Zavedenij Rossijskoj Federacii*. Sankt-Peterburg: Fakul'tet filologii i iskusstv SPbGU.

Bonfiglio, T. P. (2010). *Mother Tongues and Nations: The Invention of the Native Speaker*. Berlin: de Gruyter.

Brutt-Griffler, J., and Samimy, K. K. (2001). Transcending the nativeness paradigm. *World Engl.* 20, 99–106. doi: 10.1111/1467-971X.00199

Bunk, O. (2020). *'Aber Immer Alle Sagen Das'. The Status of V3 in German: Use, Processing and Syntactic Representation.* Ph.D. Dissertation. Berlin: Humboldt University Berlin.

Cejtlin, S. N. (2009). *Oèerki po Slovoobrazovaniju i Formoobrazovaniju v Detskoj Reèi*. Moscow: Znak.

Chatsiou, A. (2010). *A Lexical Functional Grammar Approach to Modern Greek Relative Clauses*. Ph.D. Dissertation. Colchester: University of Essex.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cook, V. (2016). "Premises of multi-competence," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and Li Wei (Cambridge: Cambridge University Press), 1–25.

Criado-Perez, C. (2019). *Invisible Women: Data Bias in a World Designed for Men*. New York, NY: Vintage Publishing.

de Saussure, F. (1916). *Cours de Linguistique Général*, eds C. Bally and A. Sechehaye (Chicago, IL: Open Court).

Embick, D., White, Y., and Tamminga, M. (2020). "Heritage languages and variation: identifying shared factors" [commentary article on Polinksy, M., and Scontras, G., "understanding heritage languages"]. *Bilingualism* 23, 21–22. doi: 10.1017/S1366728919000476

Erguvanli, E. E. (1984). *The Function of Word Order in Turkish Grammar*. Berkeley, CA: University of California Press.

Flores, C., and Rinke, E. (2020). The relevance of language-internal variation in predicting heritage language grammars. *Bilingualism* 23, 25–26. doi: 10.1017/s1366728919000464

Foth, K. A. (2006). *Eine umfassende Constraint-Dependenz-Grammatik des Deutschen*. Available online at: https://edoc.sub.uni-hamburg.de/informatik/volltexte/2014/204/ (accessed December 9, 2021).

Gagarina, N. (2002). "Productivity of the verbal inflection classes (the case of Russian)," in *Future Challenges for Natural Linguistics*, Lincom Studies in Theoretical Linguistics, Vol. 30, eds K. Dziubalska-Kolaczyk and J. Weckwerth (München: Lincom), 153–164.

Gogolin, I. (1994). *Der monolinguale Habitus der multilingualen Schule*. Münster: Waxmann.

Gogolin, I. (2002). Linguistic and cultural diversity in Europe: a challenge for educational research and practice. *Eur. Educ. Res. J.* 1, 123–138. doi: 10.2304/eerj.2002.1.1.3

Golub, I. B. (2001). *Russkij Jazyk i Kul'tura Reèi: Uèebnoe Posobie*. Moscow: Logos.

Grice, M., Baumann, S., and Benzmüller, R. (2005). "German intonation in autosegmental-metrical phonology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun (Oxford: Oxford University Press), 55–83. doi: 10.1093/acprof:oso/9780199249633.003.0003

Grosjean, F. (1982). *Life With Two Languages*. Cambridge, MA: Harvard University Press.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Guijarro-Fuentes, P., and Schmitz, K. (2015). The nature and nurture of heritage language acquisition. *Lingua* 164, 239–250. doi: 10.1016/j.lingua.2015.05.008

Hawkins, J. A. (1984). A note on referent identifiability and co-presence. *J. Pragmat.* 8, 649–659. doi: 10.1016/0378-2166(84)90004-3

Hawkins, R., Al-Eid, S., Almahboob, I., Athanasopoulos, P., Chaengchenkit, R., Hu, J., et al. (2006). Accounting for English article interpretation by L2 speakers. *EUROSLA Yearb.* 6, 7–25. doi: 10.1075/eurosla.6.04haw

Hickmann, M., and Hendriks, H. (1999). Cohesion and anaphora in children's narratives: a comparison of English, French, German, and Mandarin Chinese. *J. Child Lang.* 26, 419–452. doi: 10.1017/S0305000999003785

Holton, D., Mackridge, P., and Philippaki-Warburton, I. (1997). *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.

Igarashi, Y. (2006). "Intonation patterns in Russian interrogatives – phonetic analyses and phonological interpretations," in *Usage-Based Linguistic Informatics: 3 Prosody and Syntax*, eds Y. Kawaguchi, I. Fonagy, and T. Moriguchi (Amsterdam: John Benjamins), 175–196. doi: 10.1075/ubli.3.11iga

Ionin, T. (2006). This is definitely specific: specificity and definiteness in article systems. *Nat. Lang. Semant.* 14, 175–234. doi: 10.1007/s11050-005-5255-9

Ionin, T., and Díez-Bedmar, M. B. (2021). "Article use in Russian and Spanish learner writing at CEFR B1 and B2 Levels: effects of proficiency, native language, and specificity," in *Learner Corpus Research Meets Second Language Acquisition*, eds B. Le Brun and M. Paquot (Cambridge: Cambridge University Press), 10–38. doi: 10.1017/9781108674577.003

Ionin, T., Choi, S. H., and Liu, Q. (2021). Knowledge of indefinite articles in L2-English: online vs. offline performance. *Second Lang. Res.* 37, 121–160. doi: 10.1177/0267658319857466

Ionin, T., Zubizarreta, M. L., and Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua* 118, 554–576. doi: 10.1016/j.lingua.2006.11.012

Kornfilt, J. (1997). *Turkish. Descriptive Grammars*. Abingdon-on-Thames: Routledge.

Krause, T. (2019). *ANNIS: A Graph-Based Query System for Deeply Annotated Text Corpora.* Ph.D. Dissertation. Berlin: Humboldt-Universität zu Berlin. doi: 10.18452/19659

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

Lardiere, D., Smith, C. E., Brugos, A., and Micciulla, L. (2004). "Knowledge of definiteness despite variable article omission in second language acquisition," in *Proceedings of the 28th Annual Boston University Conference on Language Development* (Somerville, MA: Cascadilla Press), 328–339.

Leiss, E. (2010). "Koverter Abbau des Artikels im Gegenwartsdeutschen," in *Kodierungstechniken im Wandel. Das Zusammenspiel von Analytik und Synthese im Gegenwartsdeutschen*, eds D. Bittner and L. Gaeta (Berlin: de Gruyter), 137–158.

Li Wei (2016). "Epilogue: multi-competence and the translanguaging instinct," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and Li Wei (Cambridge: Cambridge University Press), 533–543.

Lithoksoou, D. (2019). *To sintaktiko Phenomeno Tis Ipotaksis sta Kimena Helinoamerikanon Heritage Omiliton (Speakers) Tis Helinikis Glosas (In Greek).* Athens: Kapodistrian University.

Loban, W. (1976). *Language Development: Kindergarten Through Grade Twelve. Research Report No 18*. Urbana, IL: National Council of Teachers of English.

Lohndal, T., Rothman, J., Kupisch, T., and Westergaard, M. (2019). Heritage language acquisition: what it reveals and why it is important for formal linguistic theories. *Lang. Linguist. Compass* 36:165. doi: 10.1111/lnc3.12357

Lowe, R. J. (2020). *Uncovering Ideology in English Language Teaching.* Berlin: Springer.

Makarova, V. (2003). "Intonation features in categorization of Russian sentence type," in *Investigations Into Formal Slavic Linguistics IV*, eds P. Kosta and J. Blaszczak (Frankfurt/M., Germany: Peter Lang), 83–95.

McCawley, J. D. (1993). "Gapping with shared operators," in *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 19*, eds J. S. Guenter, B. A. Kaiser, and C. C. Zoll (Berkeley, CA: BLS), 245–253. doi: 10.3765/bls.v19i1.1507

McGregor, A. (2020). *Sex Matters: How Male-Centric Medicine Endangers Women's Health and What We Can Do about It*. London: Quercus Publishing.

Meyer, R., and Mleinek, I. (2006). How prosody signals force and focus – a study of pitch accents in russian yes-no questions. *J. Pragmat.* 38, 1615–1635. doi: 10.1016/j.pragma.2005.05.011

Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: Benjamins. [*Studies in Bilingualism* 39]

Montrul, S., and Polinsky, M. (2011). Why not heritage speakers? *Linguist. Approaches Biling.* 1, 58–62.

Montrul, S., and Polinsky, M. (2019). "Introduction to heritage language development," in *Oxford Handbook of Language Attrition*, eds M. Schmid and B. Köpke (Oxford: Oxford University Press), 419–433.

Myers-Scotton, C. (2006). *Multiple Voices. An Introduction to Bilingualism*. Malden, MA: Blackwell.

Ortega, L. (2009). *Understanding Second Language Acquisition*. London: Hodder Education.

Ortega, L. (2016). "Multi-competence in second language acquisition: inroads into the mainstream?", in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and Li Wei (Cambridge: Cambridge University Press), 50–76. doi: 10.1017/cbo9781107425965.003

Payne, J., and Huddleston, R. (2002). "Nouns and noun phrases," in *The Cambridge Grammar of the English Language*, eds R. Huddleston and G. K. Pullum (Cambridge: Cambridge University Press), 323–524. doi: 10.1017/9781316423530.006

Polinsky, M. (2015). Heritage languages and their speakers: state of the field, challenges, perspectives for future work, and methodologies. *Z. Fremdsprachenforschung* 26, 7–27.

Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.

Pullum, G. K., and Huddleston, R. (2002). "Preliminaries," in *The Cambridge Grammar of the English Language*, eds R. Huddleston and G. K. Pullum (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781316423530.002

Rathcke, T. (2006b). "Relevance of f0 peak shape and alignment for the perception of a functional contrast in Russian," in *Proceedings of the 3rd conference on speech prosody*, (Dresden), 65–68.

Rathcke, T. (2006a). "A perceptual study on Russian questions and statements," in *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (AIPUK 37)*, eds J. Harrington, C. Mooshammer, and F. Kleber (Kiel, Germany: Institut für Phonetik und Digitale Sprachverarbeitung), 51–62.

Rathcke, T. (2009). *Komparative Phonetik und Phonologie der Intonationssysteme des Deutschen und Russischen. (Comparative Phonetics and Phonology of the Intonation Systems of German and Russian.* Univ. Dissertation. München: Herbert Utz Verlag.

Rathcke, T. (2013). On the neutralizing status of truncation in intonation: a perception study of boundary tones in German and Russian. *J. Phon.* 41, 172–185.

Riester, A., and Baumann, S. (2017). *The RefLex Scheme – Annotation Guidelines*. Stuttgart: SFB. doi: 10.18419/opus-9011

Roeper, T. (2003). "Multiple grammars, feature-attraction, pied-piping, and the question: is AGR inside TP?" in *(In)vulnerable Domains in Multilingualism*, ed. N. Müller (Amsterdam: Benjamins), 335–360.

Romaine, S. (1989). *Bilingualism*. Oxford: Basil Blackwell.

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

Sarko, G. (2009). "L2 English article production by Arabic and French speakers," in *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications*, eds M. D. P. García Mayo and R. Hawkins (Amsterdam: Benjamins), 37–66. doi: 10.1075/lald.49.06sar

Schmid, M. S., and de Leeuw, E. (2019). "Introduction to linguistic factors in language attrition," in *Oxford Handbook of Language Attrition*, eds M. Schmid and B. Köpke (Oxford: Oxford University Press), 181–190.

Schroeder, C. (1995). Post-predicate arguments in Turkish. *STUF Lang. Typol. Univ.* 48, 204–218.

Schroeder, C. (2016). "Clause combining in Turkish as a minority language in Germany," in *Exploring the Turkish Linguistic Landscape. Essays in Honour of Eser Erguvanlı-Taylan*, eds M. Güven, D. Akar, B. Öztürk, and M. Kelepir (Amsterdam: Benjamins), 81–101. doi: 10.1075/slcs.175.06sch

Schroeder, C., Iefremenko, K., and Öncu, M. (to appear). "The post-verbal position in heritage Turkish. A comparative approach with a focus on non-clausal elements," in *Zweisprachigkeit Deutsch-Türkisch. Studien in Deutschland und in den Nachbarländern*, eds Z. Kalkavan-Aydın and Y. Şimşek (Münster: Waxmann).

Seton, B., and Schmid, M. S. (2016). "Multi-competence and first language attrition," in *Cambridge Handbook of Linguistic Multi-competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 338–354.

Shadrova, A. (2020). *Measuring Coselectional Constraint in Learner Corpora: A Graph-Based Approach.* Ph.D. Dissertation. Berlin: Humboldt-Universität zu Berlin. doi: 10.18452/21606

Snape, N., García Mayo, M. D. P., and Gürel, A. (2013). L1 transfer in article selection for generic reference by Spanish, Turkish and Japanese L2 learners. *Int. J. Engl. Stud.* 13, 1–28. doi: 10.6018/ijes/2013/1/138701

te Velde, J. (2017). Temporal adverbs in the Kiezdeutsch left periphery: combining late merge with deaccentuation for V3. *Stud. Linguist.* 71, 301–336.

Tracy, R. (2002). Growing (clausal) roots: all children start out (and may remain) multilingual. *Linguistics* 40, 653–686.

Tribushinina, E., van den Bergh, H., Kilani-Schoch, M., Aksu-Koç, A., Dabašinskienė, I., Hrzica, G., et al. (2013). The role of explicit contrast in adjective acquisition: a cross-linguistic longitudinal study of adjective production in spontaneous child speech and parental input. *First Lang.* 33, 594–616.

Walkden, G. (2017). Language contact and V3 in Germanic varieties new and old. *J. Comp. Ger. Linguist.* 20, 49–81. doi: 10.1007/s10828-017-9084-2

Wiese, H. (2012). *Kiezdeutsch. Ein neuer Dialekt entsteht*. München: C.H. Beck.

Wiese, H. (2013). "What can new urban dialects tell us about internal language dynamics? The power of language diversity," in *Linguistische Berichte, Special Issue 19 [Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik]*, eds W. Abraham and E. Leiss (Hamburg: Buske), 207–245.

Wiese, H. (2020). "Language situations: a method for capturing variation within speakers' repertoires," in *Methods in Dialectology XVI*, ed. Y. Asahi (Frankfurt: Peter Lang), 105–117. [Bamberg Studies in English Linguistics].

Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., et al. (2019). *RUEG Corpus [Data set]*. Zenodo.

Wiese, H., and Müller, H. G. (2018). "The hidden life of V3: an overlooked word order variant on verb-second," in *Linguistische Berichte, Special Issue 25 [Non-Canonical Verb Positioning in Main Clauses]*, eds S. Müller and M. Antomo (Hamburg: Buske), 201–224.

Wiese, H., and Rehbein, I. (2015). "Coherence in new urban dialects: a case study," in *Lingua 172/173: [Special Issue on Coherence, Covariation and Bricolage]*, eds G. Guy and F. Hinskens 45–61. doi: 10.1016/j.lingua.2015.10.016

Wiese, H., Simon, H., Zimmer, C., and Schumann, K. (2017). "German in Namibia: a vital speech community and its multilingual dynamics," in *Language Contact in the German Colonies*: *Papua New Guinea and Beyond*, eds P. Maitz and C. A. Volker 221–245.

Zdorenko, T., and Paradis, J. (2008). The acquisition of articles in child second language English: fluctuation, transfer or both? *Second Lang. Res.* 24, 227–250. doi: 10.1177/0267658307086302

Zdorenko, T., and Paradis, J. (2012). Articles in child L2 English: when L1 and L2 acquisition meet at the interface. *First Lang.* 32, 38–62. doi: 10.1177/0142723710396797

Zemskaja, E. A. (1973). *Russkaja Razgovornaja Reè*'. Moscow: Nauka.

Zimmer, C., Wiese, H., Simon, H. J., Zappen-Thomson, M., Bracke, Y., Stuhl, B., et al. (2020). Das Korpus *Deutsch in Namibia (DNam)*: eine Ressource für die Kontakt-, Variations- und Soziolinguistik. *Dtsch. Sprach.* 48, 210–232. doi: 10.37307/j.1868-775X.2020.03.03

Zuban, Y., Rathcke, T., and Zerbian, S. (2020). "Intonation of yes-no questions by heritage speakers of Russian," in *Proceedings of 10th Speech Prosody* (Tokyo), 96–100.

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

frontiers in Psychology

SYSTEMATIC REVIEW
published: 21 February 2022
doi: 10.3389/fpsyg.2022.757023

Changes in Native Sentence Processing Related to Bilingualism: A Systematic Review and Meta-Analysis

Patricia Román[1,2]* *and Irene Gómez-Gómez*[1]

[1] Departamento de Psicología, Universidad Loyola Andalucía, Seville, Spain, [2] Loyola Behavioral Lab, Universidad Loyola Andalucía, Seville, Spain

The native language changes as a result of contact with a second language, and the pattern and degree of such change depend on a variety of factors like the bilingual experience or the linguistic level. Here, we present a systematic review and meta-analysis of works that explore variations in native sentence comprehension and production by comparing monolinguals and bilinguals. Fourteen studies in the meta-analysis provided information regarding the bilingual experience and differences at the morphosyntactic level using behavioral methods. Overall, we observed that first language processing is subject to small transformations in bilinguals that occur in sentence comprehension and production. The magnitude of the changes depended on bilingual experiences, but only length of residence in an L2 setting predicted the degree of change, where shorter length of residence was associated with larger changes. Results are discussed and related to the cognitive processes that potentially cause the transformations in the first language. The present work reveals some limitations in the field that should be addressed in future studies to better understand the mechanisms behind language attrition.

Keywords: bilingualism, sentence processing, L2 to L1 influence, linguistic variation and change, systematic review, meta-analysis

OPEN ACCESS

Edited by:
Antonella Sorace, University of Edinburgh, United Kingdom

Reviewed by:
Ehab W. Hermena, Zayed University, United Arab Emirates
Patrick Sturt, University of Edinburgh, United Kingdom

***Correspondence:**
Patricia Román
proman@uloyola.es

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

Received: *11 August 2021*
Accepted: *27 January 2022*
Published: *21 February 2022*

Citation:
Román P and Gómez-Gómez I (2022) Changes in Native Sentence Processing Related to Bilingualism: A Systematic Review and Meta-Analysis. Front. Psychol. 13:757023. doi: 10.3389/fpsyg.2022.757023

INTRODUCTION

When learning a language, individuals often rely on their native language (L1) to facilitate the acquisition of the second one (L2), but learners also find that the characteristics of their L1 interfere when they are incongruent with those in the second language (Lardiere, 2009; Libben and Titone, 2009; Macizo et al., 2010; Paolieri et al., 2010; Casaponsa et al., 2015; Peristeri et al., 2018; Contemori et al., 2019, among others). Therefore, an extensive line of work in bilingualism has been devoted to exploring L1 influences on the L2.

Despite this interest, there is also evidence that contact with a second language transforms the processing of the L1 at different linguistic levels and carries a deviation from monolinguals (the so-called attrition, Schmid, 2010). For example, bilinguals are slower than monolinguals when they name pictures in their native language (e.g., Gollan et al., 2005), and they present more tip-of-the-tongue states than monolinguals (Stasenko and Gollan, 2019) what suggests that they have reduced access to words in their L1. In addition, there are conceptual shifts where bilinguals change how they

connect words to meanings (Pavlenko, 2000, 2004; Ameel et al., 2009; Pavlenko and Malt, 2011). At the grammatical level, research has found that learning a second language may yield modifications too, including gender assignments (e.g., Kaushanskaya and Smith, 2016) and parsing preferences such as changes in the likelihood with which bilinguals attach a relative clause to a specific noun in ambiguous sentences (Dussias, 2003, 2004; Dussias and Sagarra, 2007). Finally, there are also differences in brain activity, even when behavioral performance is similar between monolinguals and bilinguals (e.g., Bice and Kroll, 2015; Román et al., 2015).

Several causes have been proposed to explain the differences between monolinguals and bilinguals in their native language. Some authors state that the reduction in accessibility to the L1 representations in bilinguals may be related to a lesser frequency of L1 use (e.g., weaker links hypothesis; Gollan et al., 2008). In fact, variables associated with a reduction in the L1 input such as L2 proficiency and immersion in a context where the L2 is spoken (Gollan et al., 2008; Monaghan et al., 2017) seem to moderate the changes observed in bilinguals across time in, for example, fluency and naming tasks (Linck et al., 2009; Baus et al., 2013).

Another explanation for the phenomenon is the transfer of L2 features (Costa and Sebastián-Gallés, 2015). Investigations show bidirectional transfer (L1 to L2 and L2 to L1) not only at the lexical but also at the grammar level, which, according to the data, seems to be less permeable to influences from the L2 (Andersen, 1982; Hicks and Domínguez, 2020). One example comes from bilinguals speaking a language that allows omission of pronouns and a language where pronouns are always present. In a study by Sorace (2000), Italian–English bilinguals increased their production of overt pronouns in Italian (a null-pronoun language), and they did so in locations that were more usual in their second language (English, an overt-pronoun language) than in their native one (for example, before the verb rather than post-verbal subject pronouns in Italian–English bilinguals).

Finally, research has extensively demonstrated that, in bilinguals, both languages are co-activated even in contexts where only one is necessary (e.g., Hatzidaki et al., 2011; Bobb et al., 2020). This co-activation involves the recruitment of capacity-limited cognitive resources and mechanisms of control aimed to avoid interference from the non-intended language that may result in differences in how bilinguals process their L1 (e.g., Titone et al., 2011; Misra et al., 2012). Several circumstances can increase a load of resources, for instance, the unbalance between languages proficiency (e.g., the unintended language being more dominant), the complexity of sentence structure, or the similarity across languages. In this regard, increased co-activation can cause facilitation (e.g., cognates, words that are similar in form and meaning) or interference (e.g., homographs, words that are similar in form differ in meaning). Also, individual differences in cognitive control and working memory (Cunnings, 2017a,b) might influence the magnitude of L1 variation[1].

Regardless of the processes implicated in how the L2 influences the native language in bilinguals, L1 attrition is not a one-way road. Although, as mentioned above, the extent of L2 contact appears to be crucial in the occurrence and depth of changes in the L1, the L2 effects on the native language do not always seem constrained to high L2 proficiency, immersion, long-term exposure, or increased frequency of L2 use. Some studies reveal changes that emerge in adult learners after limited exposure to a new language in phonetic properties (Chang, 2012), lexicon (Baus et al., 2013; Bice and Kroll, 2015), and morphosyntax (Dussias et al., 2016), and with different time-courses and degrees of affection (Sorace, 2011). Again, the rapid variations observed in the L1 suggest that the native language is not a static entity, and in fact, such shifting is not necessarily linear or incremental. For example, Chang (2012) observed that native English speakers who were novice Korean learners presented after a few weeks of exposure to the L2 a phonological drift in their L1, that is, modifications that reflect the assimilation to phonetic characteristics of a different language. The author compared this group to experienced learners enrolled in the same course (Chang, 2013), and this latter group showed a reduced drift compared to their novice peers. Besides, re-immersion in the L1 environment may reverse the changes observed (Chamorro et al., 2015; Sorace, 2020), while, in other cases, the effects of the L2 contact may persist after individuals are no longer using it, and the duration of this influence diverges depending on the linguistic property under scrutiny (Linck et al., 2009).

The production and comprehension of sentences provide a rich and informative ground to explore the influence from L2 to L1. In a sentence, lexical, morphosyntactic, and pragmatic information interact, and individuals build upon these elements to convey or understand a message. Importantly, languages differ in terms of the weight that each of the aspects mentioned above has in a sentence and the information they provide, for example, case marking in German articles reveals the role of the subsequent noun in a sentence and is absent in Spanish or English. Concerning the influence of the L2 on L1 within sentence processing, linguistic levels differ in their degree of attrition after L2 acquisition, to give an example, bilinguals and monolinguals show similar responses to gender agreement violations in their native language but not to violations in verb combinations (Bergmann et al., 2015). Moreover, sentence production is more susceptible to cross-linguistic influences than comprehension (Runnqvist et al., 2013).

Although processes behind sentence production and comprehension partly coincide (e.g., Walenski et al., 2019), language production requires retrieval from memory, selection of intended representations, and speech planning processes subject to demands that may differ from reading or listening to sentences (Daneman and Green, 1986). Sentence production provides additional information too. Authors such as Schmid et al. (2013) consider that tasks where participants freely produce a discourse (for example, by asking them to report what they saw in an image or video fragment) allow bilinguals to display their entire repertoire without restrictions. All the above makes the study of sentences more informative than single-word studies about the circumstances that bilinguals face daily but more challenging to tackle.

[1]Here the term L1 variation refers to deviations from the normative use of the L1 in monolinguals.

Studied variables that modulate the size of L1 variation in sentence processing could be divided into those related to the bilingual experience *per se*, linguistic variables, and individual differences in cognition. Within the first category, language dominance (Sanoudaki and Thierry, 2015; Kasparian et al., 2017), proficiency (Opitz, 2010; Cherciov, 2011), and frequency of use (Schmid and Dusseldorp, 2010; Kasparian and Steinhauer, 2017) have been more extensively explored. The results have led researchers to considering explanations in terms of weakened representations in L1 because of a reduced input and greater co-activation and competition between linguistic representations (Steinhauer and Kasparian, 2020). In this sense, immersion in the L2 environment represents an extreme case of exposure to L2 and limited contact with L1. For example, Dussias and Sagarra (2007) compared Spanish monolinguals, Spanish–English bilinguals with limited immersion in their L2 context, and Spanish–English bilinguals with extensive immersion while reading sentences that included a relative clause and two antecedent nouns (as in "An armed robber shot the **sister** of the **actor** who was in the balcony"). In such cases, Spanish speakers have a preference to attach the relative clause ("who was in the balcony") to the first noun ("the sister"), while native speakers of English prefer the second noun attachment ("the actor"). In their study, only bilinguals with long-term immersion in their L2 environment revealed an attachment in their L1 to the second noun, similar to native speakers of English.

Age of L2 acquisition (AoA) is another variable that has been associated with the degree of change in L1. AoA has been a matter of long debate in bilingual research under the assumption that while late bilinguals can master a second language, they hardly process morphosyntactical features the way monolinguals do (Wartenburger et al., 2003; Clahsen and Felser, 2006, but see Diependaele et al., 2011; Román et al., 2021). When considering its role in L2 to L1 influence, the question is whether some properties of the L1 become resistant to changes in late learners of L2 (Schmid and Köpke, 2017). Putting aside the case of heritage speakers (unbalanced bilinguals that learned their heritage language at home in a context where there is a dominant community language; see Benmamoun et al., 2013) that may not have fully consolidated their native language before they learn the second language, investigations addressing this matter are scarce (Karayayla and Schmid, 2019). Using semi-structured interviews, Karayayla and Schmid (2019) collected free speech samples from highly proficient Turkish–English bilinguals with an AoA that range from 7 to 34 y-o and compared the occurrence of complex syntactical forms to a group of Turkish monolinguals. They did not find a relation between the AoA and the structural complexity.

Another aspect of interest within the bilingual experience is the degree of overlap between languages. As mentioned above, the extent to which languages share properties can determine the co-activation of languages, facilitating (e.g., Domínguez, 2013) or interfering with the integration of both linguistic systems (Steinhauer and Kasparian, 2020). For example, Bernolet et al. (2007) investigated whether bilinguals were more likely to repeat a syntactical structure in a language (a passive) after they had been exposed to that same structure in another language (cross-linguistic priming) when a confederate described a picture. They had participants that spoke languages with different (English–Dutch) or similar word order (Dutch–German). While they observed priming in constructions where the word order matched, they did not when it differed. Nevertheless, research shows changes in bilinguals who speak languages with different parsing preferences (noun attachment in relative clauses; Dussias, 2003, 2004; Dussias and Sagarra, 2007).

Regarding individual differences in cognition, accounts based on limited capacities due to co-activation or susceptibility to interference (Cunnings, 2017a) predict that higher cognitive control and working memory will reduce processing difficulties in bilingual sentence parsing (but see Brothers et al., 2021). Following the same rationale, cues that facilitate successful retrieval of information will help to overcome interference and bias sentence processing, as observed in studies investigating relative clause attachment (Dussias and Sagarra, 2007) or interpretation of null/overt pronouns (Tsimpli et al., 2004) and, therefore, they might be preferred and used by bilinguals (Cunnings, 2017a,b).

So far, mixed and limited results blur the connection between the mechanisms proposed to cause attrition and the outcomes in the studies, in part due to the heterogeneity of the bilingual experience. Here, we try to put together research on sentence processing to shed some light and pave the path to future inquiries.

The Present Study

Research demonstrates that the linguistic system is not a static unit but rather a malleable and adaptive organization that integrates novel entries at different levels (de Bot et al., 2007; Kroll et al., 2014; Kaan et al., 2019) and the context of sentence processing provides a broad but complex ground to investigate how various linguistic representations interact in the bilingual mind. In the present study, we present a systematic review and meta-analysis study to explore the current evidence on how the bilingual experience changes the processing of sentences in the native language to unravel different patterns and the factors that underlie them.

METHODS

We performed a systematic review and meta-analysis following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009). The protocol of this study was previously registered at PROSPERO on May 16th, 2021 (registration number CRD42021245042).

Search Strategies

The search was conducted in Web of Science (WOS), PsycINFO, PubMed, and Scopus from database inception to March 3, 2021, and the strategy comprised keywords and text words related to bilingualism, attrition, and sentence processing and comprehension. We first piloted the search strategy in WOS and then adapted it to run across PsycINFO, PubMed, and Scopus (see **Supplementary Material**). We also screened reference lists

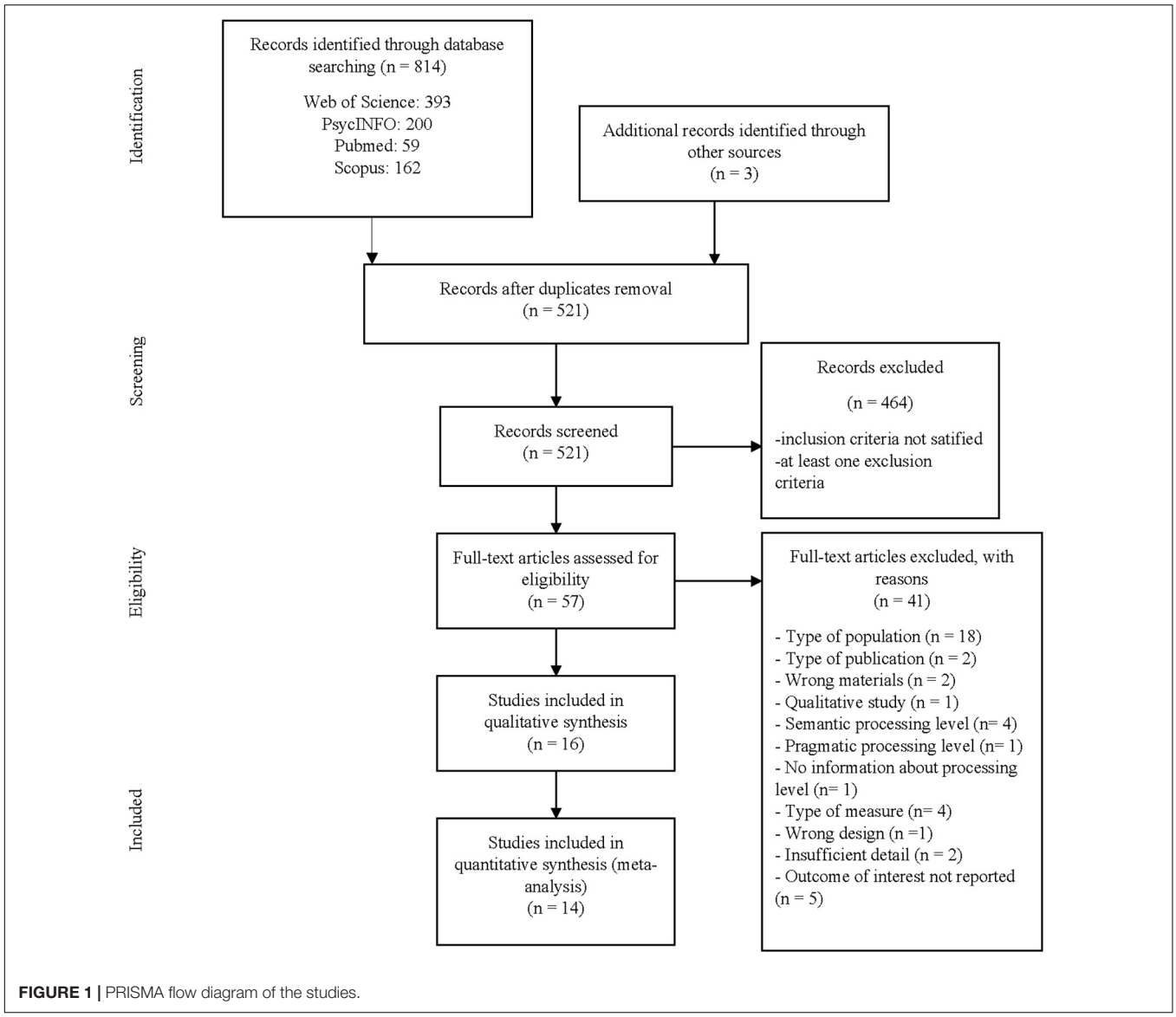

FIGURE 1 | PRISMA flow diagram of the studies.

of included studies and previous reviews on the topic. There were no language or year restrictions.

Eligibility Criteria

Following PICO criteria (population, intervention, comparator, and outcome), our inclusion criteria were as follows: the population was bilinguals, here broadly defined as participants tested in their native tongue but proficient in a second language; individuals suffering from any linguistic deficit, children, and heritage speakers were excluded; regarding intervention or exposures, we included cross-linguistic influence from the second language (L2) to the first language (L1) in morphosyntactic processing, and the comparator was morphosyntactic native language processing in bilingual participants contrasted to monolinguals' (participants with minimal to no experience with a language different from their native tongue) in both sentence comprehension and sentence production assessed by behavioral measures. Finally, the outcome of this study was the influence of the L2 on the L1, as seen through different psycholinguistic tasks and behavioral measures.

Selection of Studies

Two independent reviewers screened the title and abstract (LG and IGG) and full text to assess the eligibility of the studies. A third reviewer made the final decision (PR) in case of disagreements. The software system used for recording decisions was Rayyan (Ouzzani et al., 2015).

Data Extraction

Again, two independent reviewers (PR and IGG) extracted the data and resolved discrepancies by consensus. We collected information related to the first author, publication year, total sample size, bilingual sample size, monolingual sample size, target population (pair of languages and language experience),

cognitive processes studied (comprehension and production), task, and measures. More specifically, behavioral measures such as reaction times, accuracy, acceptability ratings in comprehension and pauses, errors, diversity, and complexity in production were included. Seven articles were discarded because of unavailable data, and information from groups of heritage speakers or children was not included in the analyses.

Statistical Analysis

Quantitative data were analyzed using the Comprehensive Meta-Analysis (CMA) software package, V.3 (CMA) (Borenstein et al., 2013) and Stata release 14.2. (StataCorp, 2015). First, standardized mean differences (SMDs) were calculated from each group's mean and SD, and when unavailable, we calculated SMDs from sample size and *p*-value. Then, CMA was used to obtain the equivalent SMDs. Finally, the pooled SMDs for all studies and their 95% CIs were estimated. Some studies only reported *p*-values corresponding to interactions or group effects that included heritage speakers groups and were discarded from the analyses.

The SMD between bilingual and monolingual participants was used as effect size using Hedges' *g* formula (Hedges, 1981) for small sample sizes. A negative value indicated higher scores in morphosyntactic tasks for the monolingual group in sentence comprehension and sentence production. Following previous evidence, the sign of the SMD was inverted for reaction time, errors, and pause measures so that the SMD went in the same direction as other measures such as accuracy in which the higher the mean, the better the performance (e.g., Beaussart et al., 2018). Interpretation of the resulting SMD followed Cohen's proposal: 0.20 as a small effect size, 0.50 as a medium effect size and, 0.80 as a large effect size (Cohen, 1988). In addition, random-effects models were used for pooling effect sizes assuming studies included heterogeneous populations that may differ from each other.

Inspection of heterogeneity was carried out through visual inspection of the forest plot, Q Cochran's statistic, and *p*-value. Additionally, heterogeneity was quantified by the I^2 index and its 95% CI, and interpretation of the I^2 index was subject to the following level and percentages: unimportant heterogeneity (0–40%), moderate heterogeneity (30–60%), substantial heterogeneity (50–90%), and considerable heterogeneity (75–100%) (Higgins and Green, 2011; Borenstein et al., 2017).

We employed the Duval and Tweedie trim-and-fill procedure (Duval and Tweedie, 2000), the Begg and Mazumdar rank correlation (Begg and Mazumdar, 1994), and the Egger test (Stuck et al., 1998) to explore publication bias.

Furthermore, sensitivity analyses were conducted to explore SMD when Cohen's *d* and the fixed-effects model were used. Finally, possible SMD variations were examined independently for studies focused on sentence comprehension, sentence production and for those exploring syntactic and morphological processing.

Finally, subgroup analyses were performed using a mixed-effects model for the categorical moderators AoA (early or late bilinguals), immersion context (either L1 or L2), length of residence (LoR) (short, long or no immersion), task modality (visual, auditorily and audiovisual), and structure congruence (similar and dissimilar across languages).

There was more than one effect size for all studies, and, therefore, the data might be considered dependent. Because of such dependence, we performed a meta-regression analysis on the continuous data with robust variance estimates (RVE; Hedges et al., 2010; Tanner-Smith and Tipton, 2014) using the *Robumeta* command in Stata. When fewer than four degrees of freedom were present, the results were considered unreliable (Tanner-Smith et al., 2016). Following previous research, covariates included in the RVE meta-regression were LoR, AoA, and L2 proficiency (e.g., Schmid and Dusseldorp, 2010). We performed both bivariate meta-regression analysis (including only one of the three independent variables in each meta-regression analysis) and multivariate meta-regression analysis (including all covariates in the same meta-regression analysis).

RESULTS

Study Selection

A total of 817 articles were identified through searching databases and other sources. After eliminating duplicates, 521 published studies and unpublished doctoral dissertations remained, and we screened the title and abstract. Fifty-seven studies initially met the inclusion criteria and were subject to a full-text inspection. This procedure yielded 16 articles that were included in the systematic review (see **Figure 1** for details about the exclusion criteria). We finally included 14 studies with 14 independent group comparisons and 81 effect sizes for the meta-analysis calculations. Two studies (Castro et al., 2017; Dragoy et al., 2019) were excluded from the meta-analysis because they did not provide the needed statistics to calculate SMD.

Study Characteristics

A total of 1,044 participants were included across 16 studies in the systematic review (see **Table 1** for study characteristics). Of them, 486 were bilinguals, and 412 were monolinguals. Ten studies included late bilinguals (studies IDs: 2, 5, 8, 9, 10, 11, 12, 14, 15, and 16), three studies included only early bilinguals (studies 3, 7, and 13), another two studies included both late, and early bilinguals (studies 1 and 6) and one study did not provide information regarding the age of acquisition (study 4). Regarding the pair of languages under scrutiny, two studies investigated Spanish–Swedish bilinguals (studies 2 and 3). The following pairs were investigated in one study each: Turkish–German (study 1), English–Spanish (study 12), Spanish–English (study 10), Italian–English (study 16), German–English (study 14), Brazilian Portuguese–European Portuguese (study 4), Chinese–Korean (study 5), Russian–German (study 6), Turkish–English (study 7), Greek–Swedish (study 8), Russian–Hebrew (study 9), German–Dutch (study 11), German–Italian (study 15), German–Spanish (study 15), Greek–English (study 16), and Spanish–Catalan (study 13). Ten studies measured sentence comprehension (studies 1, 2, 3, 4, 6, 7, 8, 9, 10, and 11) while five studies explored sentence production (studies 5,

TABLE 1 | Characteristics of the studies included in the systematic review.

			Target population						
Study ID	**References**	**Total sample size**	**Bilingual sample size**	**Monolingual sample size**	**Language experience**	**Pair of languages**	**Cognitive process**	**Task**	**Measures**
1	Arslan et al., 2015	61	LB: 20; EB: 19	22	Early and late Turkish–German bilinguals and Turkish monolinguals	Turkish–German	Sentence comprehension	Grammatical evidentiality	Behavioral (ACC and RT)
2	Bylund and Ramírez-Galan, 2014	59	39	20	Late Spanish–Swedish bilinguals and Spanish monolinguals	Spanish–Swedish	Sentence comprehension	Grammaticality Judgment test	Behavioral (AR)
3	Bylund et al., 2010	40	25	15	Early Spanish–Swedish bilinguals and Spanish monolinguals	Spanish–Swedish	Sentence comprehension	Grammaticality Judgment test	Behavioral (AR)
4	Castro et al., 2017	98	32	34	Brazilian Portuguese–European Portuguese bilinguals, Brazilian Portuguese monolinguals	Brazilian Portuguese–European Portuguese	Sentence comprehension	Acceptability judgment task	Behavioral
5	Chunpeng and Hee-Don, 2017	40	20	20	Late Chinese–Korean bilinguals and Chinese monolinguals	Chinese–Korean	Sentence production	Composition task	Behavioral (errors)
6	Dragoy et al., 2019	60	30	30	Early and late Russian–German bilinguals and Russian monolinguals	Russian–German	Sentence comprehension	Grammaticality Judgment test	Behavioral (ACC)
7	Gürel, 2015	54	27	27	Early Turkish–English bilinguals and Turkish monolinguals	Turkish–English	Sentence comprehension	Acceptability judgment task	Behavioral (AR)
8	Kaltsa et al., 2015	91	25	18	Late Greek–Swedish bilinguals and Greek monolinguals	Greek–Swedish	Sentence comprehension	Self-paced listening sentence-picture matching task	Behavioral (RT and ACC)
9	Laufer and Baladzhaeva, 2015	81	44	21	Late Russian–Hebrew bilinguals and Russian monolinguals	Russian–Hebrew	Sentence comprehension	Grammaticality judgment task	Behavioral (AR)
10	Liceras and Senn, 2009	44	12	20	Late Spanish–English bilinguals and Spanish monolinguals	Spanish–English	Sentence comprehension	Acceptability judgment task; Clitic-triggered attachment	Behavioral (AR)
11	Ribbert and Kuiken, 2010	90	52	38	Late German–Dutch bilinguals and German monolinguals	German–Dutch	Sentence comprehension	Grammaticality judgment task	Behavioral (errors)
12	Ribes and Llanes, 2015	35	15	20	Late English–Spanish bilinguals and English monolinguals	English–Spanish	Sentence and lexical production	Storytelling test; C-Cloze test	Behavioral (pauses and FO)
13	Román et al., 2015	33	16	17	Early Spanish–Catalan bilinguals and Spanish monolinguals	Spanish–Catalan	Sentence comprehension	Acceptability judgment	Behavioral (ACC)
14	Schmid, 2014	126	53	53	Late German–English bilinguals, English learners of German and German monolinguals	German–English	Sentence production	Spontaneous speech sampling	Behavioral (ACC)
15	Schmitz et al., 2016	53	I/G: 10; S/G: 8	I: 10; S: 7	Late Italian–German bilinguals, late Spanish–German bilinguals, Spanish monolinguals, and Italian monolinguals	Italian–German; Spanish–German	Sentence production	Spontaneous speech sampling	Behavioral (FO)
16	Tsimpli et al., 2004	79	I/E: 20; G/E: 19	I: 20; G: 20	Late Italian–English bilinguals, late Greek–English bilinguals, Italian monolinguals, and Greek monolinguals	Italian–English/Greek–English	Sentence production and comprehension	Headlines task; picture verification task	Behavioral (FO)

Groups within studies that do not meet the criteria are not included in the table.
Bilingual sample size: E/R, English and Russian speakers; R/E, Russian and English speaker; LB, late bilingual; EB, early bilinguals; I/G, Italian and German speakers; S/G, Spanish and German speakers; I/E, Italian and English speakers; G/E, Greek and English speakers.
Monolingual sample size: E, English speakers; R, Russian speakers; I, Italian speakers; S, Spanish speakers; G, Greek speakers.
Measures: RT, reaction times; ACC, accuracy; AR, acceptability ratings; FO, frequency of occurrence.

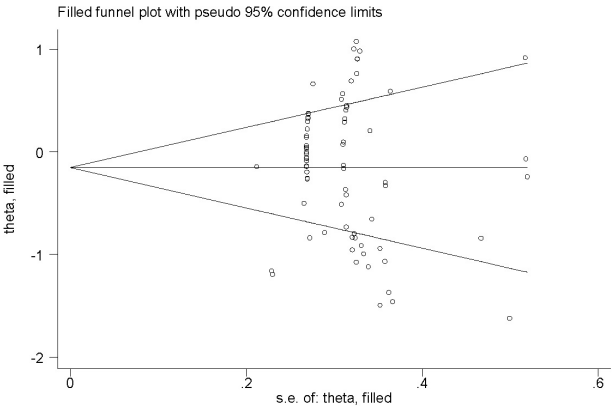

FIGURE 2 | Funnel plot.

12, 14 and 15). One study collected data from both sentence production and comprehension (study 16). All studies used behavioral measures.

Differences Between Bilinguals and Monolinguals in L1

The meta-analysis was calculated based on 81 effect sizes reported in 14 articles. The pooled SMD was −0.155 (95% CI, −0.301 to −0.009; $p < 0.001$). There was substantial heterogeneity across studies (I^2 = 79.1%; 95% CI, 74% to 83%), and it reached significance (Q_{80} = 382.99; $p < 0.001$). These results showed a statistically significant difference between monolinguals and bilinguals in L1 morphosyntactic processing, and according to Cohen's proposal, the effect size was small.

Publication Bias

Results of the Egger's test (bias, −0.907; 95% CI, −4.082 to 2.268; $p = 0.571$) and Begg and Mazumdar's test ($z = 1.02$; $p < 0.314$) indicated no publication bias. The Duval and Tweedie procedure did not impute any missing studies, and the SMD did not change [SMD, −0.155; (95% CI, −0.301 to −0.009); $p < 0.038$]. As a result, there is no evidence of publication bias. The funnel plot is shown in **Figure 2**.

Sensitivity Analysis

The pooled SMD based on the 81 effect sizes revealed little change when Cohen's *d* [SMD, −0.157 (95% CI, −0.305 to −0.008; $p = 0.039$)] and the fixed-effects model [SMD, −0.144 (95% CI, −0.210 to −0.078; $p = 0.000$)] were used. The effect size remains small in both cases, and the differences between monolingual and bilingual speakers remain statistically significant. If studies on morphological processing were independently subject to analyses, the SMD increased considerably, and the differences between monolingual and bilingual speakers remain statistically significant [SMD, −0.882 (95% CI, −1.167 to −0.598; $p < 0.001$)]. However, when studies on syntactic processing [SMD, −0.011 (95% CI, −0.155 to 0.133; $p = 0.879$)], comprehension tasks [SMD, −0.165 (95% CI, −0.339 to −0.008; $p = 0.061$)], and production tasks [SMD, −0.124 (95% CI, −0.397 to 0.148; $p = 0.372$)] were analyzed separately, the SMD varied with no significant differences between monolingual and bilingual groups. **Table 2** summarizes the results of the sensitivity analysis.

Subgroup Analysis

Table 3 depicts the results for the subgroup analysis. The studies that did not report information about some of the variables of interest (AoA, immersion context, immersion duration, modality of presentation, structure similarity) comprised a no-information group included in the corresponding subgroup analysis to explore their role in the outcome.

We inspected subgroups associated with characteristics of bilingual experience. There were statistically significant differences between subgroups investigating early, late bilinguals, and a no-information group ($Q_2 = 18.05$, $p < 0.001$) with greater effect sizes observed in studies without information related to AoA ($k = 11$), while studies including late AoA ($k = 45$) presented a greater effect compared to those comprised of bilinguals with early acquisition of the L2 ($k = 25$). However, none of the individual effects in each group was significant (all p's > 0.05). When considering immersion, no differences appeared between articles exploring bilinguals immersed in their L1 ($k = 1$) and those exploring bilinguals immersed in their L2 ($k = 80$), but this last group of studies presented a significant effect size ($p = 0.035$). Finally, when looking at LoR in the L2 environment, we found statistically significant differences between short and long LoR studies ($Q_1 = 17.76$, $p < 0.001$), and the effect was significant in the short LoR ($k = 2$), but it did not reach significance in the long LoR subgroup ($k = 79$; $p = 0.080$).

Exploration of the similarity of the structures between languages revealed that effect sizes significantly differed when we compared studies using similar and dissimilar features between languages (similar, dissimilar, collapsed, n/a, and no-information, $Q_4 = 109.21$, $p < 0.001$), but differences between monolinguals and bilinguals were not significant in the group of studies that used either similar or dissimilar structures. As seen in **Table 3**, effect sizes were larger if the languages' characteristics were collapsed ($k = 1$).

Investigating the modality of stimuli presentation, research dealing with auditory, visual, and audiovisual material did not significantly differ between modalities, and the effect size was statistically significant only in the group with visual presentation ($k = 42$; $p = 0.043$).

Meta-Regression

Meta-regression analyses with RVE (Hedges et al., 2010; Tanner-Smith and Tipton, 2014) were performed. We considered the effect of each continuous moderator individually (LoR, L2 proficiency, or L2-AoA; bivariate meta-regression) over the SMD, and the results indicated a significant difference in the effect size magnitude associated with LoR (β = 0.040 [95% CI, 0.013 to 0.066]; $p < 0.016$). However, according to Tanner-Smith et al. (2016), the *p*-value was untrustworthy because the degrees of freedom were less than four. No significant differences were observed for proficiency in L2 and L2-AoA variables (see **Table 4**).

TABLE 2 | Sensitivity analysis.

Analysis	No. of effect sizes *(k)*	SMD	95% CI	*p*	I^2 (95% CI)
Effectiveness	81	−0.155	−0.301 to −0.009	0.038	79.1% (74%–83%)
Cohen's *d* test	81	−0.157	−0.305 to −0.008	0.039	79.02% (74%–83%)
Fixed effect model	81	−0.144	−0.210 to −0.078	<0.001	79.1% (74%–83%)
Including comprehension task studies only	59	−0.165	−0.339 to 0.008	0.061	81.1% (76%–85%)
Including production task studies only	22	−0.124	−0.397 to 0.148	0.372	72.1% (57%–82%)
Including syntactic processing studies only	68	−0.011	−0.155 to 0.133	0.879	73.8% (67%–79%)
Including morphological processing studies only	13	−0.882	−1.167 to −0.598	<0.001	67.8% (43%–82%)

TABLE 3 | Subgroup analysis.

Subgroup analysis	No. of effect sizes *(k)*	SMD	95% CI	$p^{\dagger}$	I^2	Between-group heterogeneity[‡]
Type of task						
Comprehension	59	−0.165	−0.339 to 0.008	0.061	81.1%	$Q_1 = 0.46$, $p = 0.500$
Production	22	−0.124	−0.397 to 0.148	0.372	72.1%	
AoA (type of bilingualism)						
Early bilingualism	25	0.022	−0.103 to 0.147	0.734	26.2%	$Q_2 = 18.05$, $p < 0.001$
Late bilingualism	45	−0.193	−0.424 to 0.038	0.102	83.5%	
No information	11	−0.423	−0.932 to 0.086	0.104	84.9%	
Context						
Immersed in L1	1	0.206	−0.462 to 0.874	0.545	–	$Q_1 = 1.06$, $p = 0.302$
Immersed in L2	80	−0.159	−0.307 to −0.012	0.035	79.3%	
LoR						
Short	2	−1.148	−1.798 to −0.498	0.001	48.6%	$Q_1 = 17.76$, $p < 0.001$
Long	79	−0.130	−0.276 to 0.016	0.080	78.5%	
Modality						
Auditory	23	−0.202	−0.569 to 0.166	0.282	85.4%	$Q_2 = 5.32$, $p = 0.070$
Visual	42	−0.189	−0.372 to −0.006	0.043	77.7%	
Audiovisual	16	0.005	−0.262 to 0.272	0.970	65.5%	
Structure congruence						
Different	43	−0.134	−0.294 to 0.027	0.103	67.8%	$Q_4 = 109.21$, $p < 0.001$
Similar	25	0.190	−0.053 to 0.434	0.125	75.8%	
Not applicable	5	−1.054	−1.405 to −0.703	0.001	21.0%	
Collapsed	1	−1.163	−1.611 to −0.715	0.001	–	
No information	7	−0.751	−1.348 to −0.154	0.014	84.7%	

[†]*Significance tests in which for each subgroup, the null hypothesis is that SMD = 0.*
[‡]*Q-values represent the comparison of subgroup means based on a chi-square distribution in which the null hypothesis is that the effect size is the same for all subgroups.*

When the effect of all covariates (LoR and L2-AoA) were considered simultaneously (multivariate meta-regression) in the same RVE meta-regression model, LoR and age of acquisition did not predict the effect size magnitude. L2 proficiency had to be excluded from the analysis due to missing values.

DISCUSSION

The work presented here addresses the changes that the native language undergoes as a consequence of contact with a second language in bilinguals. Our aim was twofold: we wanted to explore circumstances that lead to variations in the L1 and find a connection between the explanatory accounts of bilingual/monolingual differences in native processing and the data. To do so, we employed a systematic review and meta-analysis of research in sentence comprehension and production comparing monolingual and bilingual performance. Next, we summarize our results, and then we will try to link them with cognitive processes that might be behind changes in the native processing of sentences.

The systematic review comprised 16 studies, and 14 were used in the meta-analysis. Overall, results showed that individuals who speak more than one language were subject to variations in their native tongue with a small and significant effect size. As we strove to tell apart L1 changes connected to the main facets of bilingual sentence processing, subgroups analyses were used. No differences were observed between comprehension and production, and importantly, if analyses considered each group of studies separately, monolinguals and bilinguals did not display different behavior. Regarding the bilingual experience, differences were statistically significant in studies that compared monolinguals to bilinguals immersed in their L2 and bilinguals with a short LoR in a context

TABLE 4 | Coefficient statistics of meta-regression analysis with RVE estimates on the association between SMD and other covariates.

	Beta	95% CI	*p*
Bivariate meta-regression			
LoR	0.040	0.013 to 0.066	0.016
Proficiency in L2	0.050	−0.474 to 0.574	0.730
AoA	−0.001	−0.050 to 0.050	0.964
Multivariate meta-regression			
LoR	−0.031	−0.046 to 0.108	0.254
AoA	0.019	−0.064 to 0.102	0.499

LoR, length of residence; AoA, age of acquisition.

where the second language is spoken. In the meta-analysis regression, the continuous variables L2 proficiency and AoA did not predict the effect size observed in sentence processing; following the subgroups analysis, shorter LoR predicted wider differences between monolinguals and bilinguals in their L1. All in all, these findings suggest that L1 processing may be subject to small, qualitatively different variations across the bilingual experience rather than accumulative (Schmid and Karayayla, 2020; Steinhauer and Kasparian, 2020), and all the factors should be taken into consideration when addressing the patterns of bilingual processing in their native language.

Comprehension and Production Studies

Sentence production is less explored than comprehension despite the stress that recent approaches give to considering both production and comprehension and its interdependence to feed language models (Pickering and Garrod, 2013; Dell and Chang, 2014). Our work reflects this trend in research, with a higher number of studies targeting comprehension.

Within comprehension, most of the reviewed studies used acceptability/grammaticality judgments (12 studies; 10 included in the meta-analysis, see **Table 1**). In an investigation by Schmid and Dusseldorp (2010), German monolinguals outperformed German–English bilinguals and German–Dutch bilinguals in verbal fluency, the C-test, and a film re-telling task; however, their behavior was similar in an auditorily grammatical judgment task. Although their results may reveal stable L1 knowledge at this linguistic skill in bilinguals, the authors do not discard the hypothesis that the sentences could be easy for the readers. Also, the use of brain markers as participants read sentences may reveal differences that the offline acceptability judgments do not, even in low complexity sentences (see Román et al., 2015). For example, Italian–English bilinguals and Italian monolinguals in Kasparian et al. (2017) had to rate the grammaticality of eight-word sentences in Italian that could include local or non-local agreement violations while neurophysiological activity was recorded (for example, "Il lavatore torna dalla fabbrica sporco di grasso," *The* ***workers***$_{plural}$ *returns*$_{singular}$ *from the factory dirty*$_{singular}$ *with grease*). They did not find a group effect or group interaction at the acceptability ratings, but groups differed in reaction times and the neurophysiological patterns. Together with neurophysiological methodologies, other online measures like eye-tracking or self-paced reading may be more sensitive than offline acceptability judgments to catch differences between groups (e.g., self-paced listening, Kaltsa et al., 2015, in our review; for a deeper discussion on online measures in the field, see Marinis, 2003; Roberts, 2012). The prevalence of the acceptability-ratings task and behavioral measures within our sample of studies might be behind the lack of a subgroup effect.

Research dealing with sentence production, on the other hand, employed spontaneous speech/writing sampling (speech, Schmid, 2014; Schmitz et al., 2016; writing, Chunpeng and Hee-Don, 2017), storytelling (Ribes and Llanes, 2015), and a headlines task, which provides a verb, a noun phrase and an adverbial expression that participants have to use to produce a sentence describing a picture (Tsimpli et al., 2004). When present, differences in these studies showed that bilinguals tended to be slower and employed syntactical structures that were permitted in their L1 but preferred in their L2, suggesting that variation, rather than attrition or loss is a term better suited for L1 changes in individuals that speak more than one language (Schmid, 2014; Schmitz et al., 2016). Although a separate analysis of bilinguals vs. monolinguals in sentence production studies did not show a significant effect size, the production pattern reveals co-activation and transfer of L2 features in the bilingual grammar, as we will explain later.

Bilingual Experience

The magnitude of effect sizes differed between groups within AoA, and LoR, as seen in the subgroups analysis, but were only statistically significant when looking at studies that included immersion in L2 and short LoR. Nevertheless, LoR was the only continuous variable predicting differences between monolinguals and bilinguals in the bivariate meta-regression. In Schmid and Dusseldorp's (2010) study mentioned above, one of their objectives was investigating variables that predicted L1 attrition. Only LoR predicted attrition in free speech (in lexical diversity and errors) using a film re-telling task. Some articles in the present review that did not investigate the relation of attrition with LoR used a similar procedure. For example, Schmid (2014) had a subgroup of 20 late German/English bilinguals with a minimum LoR in Canada of 9 years but up to four decades (mean and standard deviation not provided) that did not differ from their monolingual peers in morphosyntactic variables. In our meta-regression, lower LoR is associated with larger differences between bilinguals and monolinguals and grants the need for further research, including an earlier (and shorter) range of LoR in sentence production.

More recently, Schmid and Karayayla (2020) explored the role of LoR in comprehension. They collected sentence production data from 92 Turkish-English bilinguals (collapsed including heritage speakers and therefore excluded from our review and meta-analysis) about their L1 maintenance and acquisition within a wide range of age at onset of bilingualism (AaO; from birth to adulthood). The data indicated that LoR predicted morphosyntactic complexity in L1, but in a direction opposite to our results, the longer the residence in a context where the second language is used, the lower the proficiency in L1. Importantly, they observed that the effects were more evident in early than late bilinguals. Because in the present work 11 of the 14 studies included late bilinguals, it is necessary to assume that variations

in the L1 are qualitatively different between bilingual experiences and likely the result of different cognitive processes.

In spite of that, the continuous variables LoR, L2 proficiency, and AoA did not predict variation in the L1 in the multivariate meta-regression. One potential explanation is that bilinguals can use cues of different nature to compensate for differences, masking the effect of the variables of interest at the behavioral level examined here. For example, in a study collecting neurophysiological data, Kasparian and Steinhauer (2017) used an acceptability judgment task in Italian with relative clause structures that could be temporarily ambiguous (garden-path as in "Il poliziotto che i ladri arresta registra i nomi," *The policeman that the thieves* ***arrests*** *registers the names*). The sentences were grammatical in Italian, but some were ungrammatical in English (as in the example) and less preferred garden-path structures in Italian. As expected, bilinguals found the grammatical sentences in Italian but ungrammatical in English as less acceptable than monolinguals. Moreover, the authors predicted a P600 in both groups (greater in bilinguals) to the verb in the relative clause (*arresta*), commonly found in garden-path sentences as an index of syntactic difficulty, and an N400 related to difficulties in semantic integration because they introduced strong semantic cues (policeman-thieve-arrest) that did not conflict with the structure. Italian monolinguals, which rely more on semantic information, evinced an N400, while the bilinguals did not show this component but a greater P600, as anticipated if they used, like English monolinguals, the strict word order preferably than semantic cues.

One problem in our study that may prevent us from finding a stronger impact of L2 proficiency is using the bilingual term and the collection of L2 proficiency data broadly to cover as many studies as possible. Some authors have warned about the implications that the way we conceive, and measure bilingualism have on the diversity of outcomes we obtain in our growing field of knowledge (Green and Abutalebi, 2013; Luk and Bialystok, 2013; Ooi et al., 2018; Anderson et al., 2020; Kremin and Byers-Heinlein, 2021). In the pool of articles reviewed here, proficiency tests go from subjective reports, including those with questions about how much effort it takes to use a language (Liceras and Senn, 2009), to objective measures and placement tests (for example, TOEFL for English proficiency, Chang, 2009; Gürel, 2015; telc for Turkish, Arslan et al., 2015). In addition, some works did not contain proficiency information (e.g., Kaltsa et al., 2015), or provided ranges (e.g., Bylund and Ramírez-Galan, 2014). Under such circumstances, there was enough data to consider proficiency as a categorical variable but impeded the continuous data to be used as recommended for the meta-regression (Bialystok, 2018; De Cat et al., 2018; Gunnerud et al., 2020).

Connection to Explanatory Accounts

As mentioned in the introduction, researchers have considered three leading causes behind the patterns observed in bilingual L1 processing: language co-activation in bilinguals, cross-linguistic transfer, and a reduced frequency of L1 use (Costa and Sebastián-Gallés, 2015; Cunnings, 2017a; Schmid and Köpke, 2017).

Cunnings (2017a) proposes that retrieval interference/facilitation lies behind differences in bilingual comprehension. Retrieval interference appears because working memory is a capacity-limited entity (Baddeley, 2013), and co-activation of languages increases the demands when bilinguals have to select one representation from those retrieved from long-term memory and integrate it with the incoming information. In agreement with this idea, Chunpeng and Hee-Don (2017) tested Chinese/Korean bilinguals reporting the use of both languages daily, and therefore more prone to co-activation. Their sample presented difficulties retrieving words in written composition and spent a longer time completing the task compared to Chinese monolinguals. Apart from that, in their research, grammar differences were mainly related to transfer from their L2 (Korean word order, punctuation, among others), similar to English/Spanish bilinguals in Ribes and Llanes (2015) that produced more subordinate constructions allowed in English but preferred in Spanish, and more pauses. Bilinguals may then become "opportunistic" speakers under co-activation contexts selecting representations that alleviate their cognitive load and languages cooperate rather than interfere (Beatty-Martínez et al., 2020).

A prediction derived from the above is that similarities across languages may influence bilingual processing, increasing co-activation, interfering or facilitating processing, or making bilinguals opting for the common structures in sentence processing. However, we have not observed differences between groups with similar and dissimilar structures, and none of the analyses that independently tested similarity and dissimilarity in bilingual vs. monolingual performance was significant. In this respect, co-activation might happen not only when the use of specific structures sharing properties across languages spread activation to nodes in the L2, but also when using the L1 in broader contexts such as L2 immersion (as seen confirmed in the meta-analysis) or when co-activation is locally induced (e.g., watching a movie in the L2; Elston-Güttler et al., 2005). This indicates that any context that prompts co-activation may increase the chances of altered processing even in the absence of shared properties.

In association with this idea, Green and Abutalebi (2013) proposed the adaptive control hypothesis. According to this approach, bilinguals find themselves in different contexts, and each posits specific demands to which the reader/speaker adapts. That is, bilinguals' L1 (and L2) is not only subject to differences in the long run but varies depending on the demands of the environment (e.g., the accent of the interlocutor, unilingual workplace, bilingual community, discourse complexity, etc.). Whenever co-activation creates competition between language schemas, there will be a need for processes that handle interference and select the desired linguistic representations (Green and Abutalebi, 2013). Such processes, on the one hand, use limited resources that may be unavailable for efficient processing (Perfetti and Stafura, 2014; Li and Clariana, 2019) and, on the other hand, make competing representations less accessible through inhibition (e.g., Levy et al., 2007). These processes engaged in controlling interference might explain the direction of LoR effects in our data; at short LoR, the native language is dominant in late bilinguals and L2 usage is expected to trigger inhibitory processes to overcome the interference and facilitate retrieval of the weaker L2 representations. When

bilinguals try to retrieve their L1 later, it takes time to access the suppressed representations in the L1. As asymmetry between languages decreases, inhibition is no longer needed (Levy et al., 2007).

Also, predictions from this approach are that reading and producing sentences will show either an impairment (for example, an increasing number of pauses, reduced processing speed) or compensation (for example, using simpler structures, using a limited number of cues, repetition of more accessible structures). As discussed earlier, bilinguals in some of the studies reviewed here present more pauses and longer time to produce written or verbally sentences in their L1 (Ribes and Llanes, 2015; Chunpeng and Hee-Don, 2017). Additionally, it is expected that bilinguals (and monolinguals) will have to overcome co-activation and inhibition, and that succeeding in doing so must be more difficult in constructions that demand resources allocation, such as complex grammatical forms (e.g., evidentiality marking in Turkish, Arslan et al., 2015) or pronominal referential relations in sentence comprehension (Li and Clariana, 2019). In fact, pronominal resolutions have been studied in the context of language attrition, considering that difficulties in this type of structure may arise due to the need to coordinate different interfaces (grammatical and discourse information; Sorace, 2011; Chamorro et al., 2015). Giving support to this view, we have found research in which bilinguals do not have a strong preference toward a pronominal assignment in the presence of more than one antecedent noun as monolinguals have (Kaltsa et al., 2015; Castro et al., 2017; but see Liceras and Senn, 2009; Schmitz et al., 2016) and show preferences that place less cognitive demands on the reader (low-attachment in L1 in both English–Spanish and Spanish–English bilinguals, Dussias, 2003, 2004), thereby supporting the notion of bilinguals as strategists that adapt to the requirements of their linguistic and cognitive context.

Another factor that may alter L1 processing is the cross-linguistic transfer (Costa and Sebastián-Gallés, 2015). In this regard, we have seen that bilinguals show in their L1 a bias to structures and interpretations that appear frequently in their L2 (examples in production are Ribes and Llanes, 2015; Chunpeng and Hee-Don, 2017; in comprehension, Dussias and Sagarra, 2007). Changes in the L1 sentence processing can appear as a lack of preference for both L1 and L2 biases too, similar to what Kaltsa et al. (2015) and Castro et al. (2017) observed. Specifically, in two experiments, Kaltsa et al. (2015) employed a self-paced listening task with a sentence-picture matching to test pronoun resolution in Greek. They presented sentences such as "I γiaγia xeretise tin kopela otan afti pernuse to δromo" (*The old lady greeted the girl when she crossed the street*) where the antecedent of the overt pronoun in the subordinate clause "afti" (*she*) could be the subject (old lady) or the direct object (the girl). In Greek, the null subject is the default, and the subject antecedent is preferred under such conditions, while a non-subject antecedent (the girl in the example) is preferred in overt subject pronoun as the one in the example. In this research, Greek–Swedish bilinguals with long LoR were expected to choose more subject antecedents in overt pronoun conditions than monolinguals, but their performance should be similar in null pronoun conditions (Swedish is a non-null subject language). Monolinguals showed differences between null and overt pronoun conditions, but bilinguals did not. Despite this outcome, Kaltsa and colleagues consider that cross-linguistic transfer is not causing their results because a group of older monolinguals presented a performance closer to bilinguals.

Finally, a frequency-based account highlights a reduced L1 input that weakens and biases activation toward L2 properties (Gollan et al., 2005). One way to explore this effect would have been to include relative frequency of use across languages in our analyses, but only a few had information about it (Liceras and Senn, 2009; Castro et al., 2017), and they did not explore its impact. Immersion length may switch the frequency balance between L1 and L2, and studies exploring parsing preferences that differ across languages are helpful since language biases may progressively shift with exposure. However, the results in our meta-regression with shorter LoR showing larger effects say otherwise. In two studies, Spanish–English bilinguals with long L2 immersion (Dussias and Sagarra, 2007) and Spanish–English bilinguals with shorter L2 immersion (Dussias, 2004) but similar proficiency change from L1 to L2 preferences while highly proficient Spanish–English bilinguals immersed in their L1 maintained the monolingual routines (Dussias and Sagarra, 2007). Such results could suggest that frequency-related modifications occur rapidly and appear in readers with shorter exposure to L2 but also that there is more than one mechanism underlying variations across bilingual groups in agreement with evidence that shows a non-linear L1 variation (Schmid and Karayayla, 2020). Other processes such as the need for L1 inhibition in an L2 setting cannot be ruled out, and ERPs studies point to a more significant role of competition in grammar differences instead of a reduced frequency in L1, at least in the early stages of immersion in an L2 context (for a review, see Steinhauer and Kasparian, 2020).

Although these explanations do not exclude each other and may act simultaneously, they do predict different outcomes under different conditions. Therefore, more research is necessary to separate their respective effects and the relative weight they have on bilinguals' L1 at several stages and situations.

CONCLUSION

To sum up, learning an L2 involves mutual influence between languages that vary qualitatively and quantitatively across time and experience, most likely in response to differences in demands of the environment and the cognitive processes recruited to deal with them.

In the present work, we targeted morphosyntactic processing. Some researchers consider that attrition occurs mainly at the lexical level in comprehension and production and that when grammatical rules are concerned, it affects aspects related to lexical retrieval (Schmid and Fägersten, 2010). Nonetheless, we have observed that bilingual deviations from monolinguals are evident though small in morphosyntax, even in individuals immersed in their second language but without knowledge in their L2 (Laufer and Baladzhaeva, 2015). It is important to note that studies targeting morphosyntax often imply the interplay

of distinct linguistic subskills, as seen above in Kasparian and Steinhauer (2017), where semantic constraints influence bilingual parsing of garden-path sentences. Experimental designs that dissociate effects across linguistic levels will help to clarify differences in the processes subserving L2 influences in the L1.

Other variables not addressed here may influence the cognitive processes involved in sentence processing, resulting in differences between monolinguals and bilinguals in their native language. The two reasons behind this absence are the lack of consistency in gathering such data among studies and selecting contrasts between monolinguals and bilinguals instead of different groups within bilinguals. Further research will need to attend to these variables. For example, none of the experiments reviewed collected data about individual differences in either cognitive control (conflict detection, inhibition, cognitive flexibility, or task switching) or working memory, which impeded directly testing the association of cognitive resources in dealing with language co-activation. Besides, here the target contrast was comparing bilinguals to monolinguals. Hopp and Schmid (2013) warned of the difficulties that using monolinguals as a reference group entails, given that the mere existence of two languages in the bilingual mind provides a qualitatively different ground. In this sense, it seems more appropriate to compare early and late bilinguals who reach similar proficiency or investigate individual differences while matching bilingual experience. While we acknowledge the validity of Hopp and Schmid's statement, we consider that there is a value in our work to address the distinct mechanisms involved in dealing with two languages and their relation with several aspects of the bilingual experience.

Apart from the above, the results presented require cautious interpretation. Although data dependence was controlled in the meta-regression analysis with RVE (Hedges et al., 2010; Tanner-Smith and Tipton, 2014), subgroups analyses, sensitivity, and main effect analyses did not control it. Besides that, we did not assess the risk of bias of the included studies. Aside from the methodological limitations, reported/collected information regarding the bilingual experience was quite heterogeneous across studies what prevented us from having a more extensive and detailed exploration of factors impacting language processing.

Despite these limitations, the present work followed the PRISMA guidelines; thus, the search of the studies was carried out in the most relevant databases using a wide variety of terms without restrictions regarding the year of publication or language, giving the systematic review and meta-analysis a high sensitivity. Additionally, two independent researchers went through the entire screening process, selection, and extraction of the characteristics. Last but not least, there was no evidence of publication bias, and we analyzed sensitivity, which contributes to the robustness of the results found.

Finally, some recommendations can be derived from our work. Methodologically speaking, offline measures, when used, should be accompanied by online measures that provide more sensitivity and information regarding the cognitive processes involved. In addition, because sentence processing implies the interaction of several types of information, the inclusion of conditions that allow isolating the influence of independent variables is encouraged. Regarding the target population, the study of cross-linguistic influence requires (1) a clear definition of terms like "proficiency"; (2) detailed data collection of measures and experience with both languages to have a profile that considers the balance between L1 and L2, and (3) going beyond monolinguals as a group of reference and consider bilingual groups that differ in several dimensions to explore the variables that affect changes in the L1.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

PR conceived the study. PR and IG-G prepared the strategy search, conducted the systematic literature search, screened studies for eligibility together with Laura Giraldo and Raquel Román, extracted data, contributed to the manuscript preparation, and approved the final version. IG-G conducted the statistical analyses. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was financially supported by grants from the Spanish Ministry of Science, Innovation, and Universities (Fondos FEDER) and the Andalusian Government (PGC2018-093506-B-I00 and PY18-FR-0007).

ACKNOWLEDGMENTS

The authors would like to thank Laura Giraldo and Raquel Román for their assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.757023/full#supplementary-material

REFERENCES

Ameel, E., Malt, B. C., Storms, G., and van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *J. Mem. Lang.* 60, 270–290.

Andersen, R. W. (1982). "Determining the linguistic attributes of language attrition," in *The loss of language skills*, eds R. D. Lambert and B. F. Freed (Rowley, MA: Newbury House), 83–118.

Anderson, J., Hawrylewicz, K., and Bialystok, E. (2020). Who is bilingual? Snapshots across the lifespan. *Bilingual. Lang. Cognit.* 23, 929–937. doi: 10.1017/S1366728918000950

Arslan, S., Bastiaanse, R., and Felser, C. (2015). Looking at the evidence in visual world: eye movements reveal how bilingual and monolingual Turkish speakers process grammatical evidentiality. *Front. Psychol.* 6:1–13. doi: 10.3389/fpsyg.2015.01387

Baddeley, A. D. (2013). *Essentials of human memory* (classic ed.). London: Psychology Press.

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

Beatty-Martínez, A. L., Navarro-Torres, C. A., and Dussias, P. E. (2020). Codeswitching: A Bilingual Toolkit for Opportunistic Speech Planning. *Front. Psychol.* 11:1699. doi: 10.3389/fpsyg.2020.01699

Beaussart, M. L., Barbarot, S., Mauger, C., and Roy, A. (2018). Systematic review and meta-analysis of executive functions in preschool and school-age children with neurofibromatosis Type 1. *J. Int. Neuropsychol. Soc. JINS* 24, 977–994. doi: 10.1017/S1355617718000383

Begg, C. B., and Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics* 50:1088. doi: 10.2307/2533446

Benmamoun, E., Montrul, S., and Polinsky, M. (2013). Defining an "ideal" heritage speaker: theoretical and methodological challenges Reply to peer commentaries. *Theoret. Linguist.* 39, 259–294. doi: 10.1515/tl-2013-0018

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

Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2007). Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *J. Exp. Psychol. Learn. Mem. Cognit.* 33, 931–949. doi: 10.1037/0278-7393.33.5.931

Bialystok, E. (2018). "Bilingualism and executive functions: What is the connection?," in *Bilingual cognition and language*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins Publishing Company), 283–306. doi: 10.1075/sibil.54.13bia

Bice, K., and Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport* 26, 966–971. doi: 10.1097/WNR.0000000000000453

Bobb, S. C., Von Holzen, K., Mayor, J., Mani, N., and Carreiras, M. (2020). Co-activation of the L2 during L1 auditory processing: An ERP cross-modal priming study. *Brain Lang.* 203:104739. doi: 10.1016/j.bandl.2019.104739

Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2013). *Comprehensive Meta-Analysis Version 3*. Englewood: Biostat.

Borenstein, Michael, Higgins, J. P. T., Hedges, L. V., and Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Res. Synth. Methods* 8, 5–18. doi: 10.1002/jrsm.1230

Brothers, T., Hoversten, L., and Traxler, M. (2021). Bilinguals on the garden-path: Individual differences in syntactic ambiguity resolution. *Bilingual. Lang. Cognit.* 24, 612–627. doi: 10.1017/S1366728920000711

Bylund, E., Abrahamsson, N., and Hyltenstam, K. (2010). The role of language aptitude in first language attrition: The case of pre-pubescent attriters. *Appl. Linguist.* 31, 443–464. doi: 10.1093/applin/amp059

Bylund, E., and Ramírez-Galan, P. (2014). Language Aptitude in First Language Attrition: A Study on Late Spanish-Swedish Bilinguals. *Appl. Linguist.* 37, 621–638. doi: 10.1093/applin/amu055

Casaponsa, A., Antón, E., Pérez, A., and Duñabeitia, J. A. (2015). Foreign language comprehension achievement: insights from the cognate facilitation effect. *Front. Psychol.* 6:588. doi: 10.3389/fpsyg.2015.00588

Castro, T., Rothman, J., and Westergaard, M. (2017). On the directionality of cross-linguistic effects in bidialectal bilingualism. *Front. Psychol.* 8:1–14. doi: 10.3389/fpsyg.2017.01382

Chamorro, G., Sorace, A., and Sturt, P. (2015). What is the Source of L1 Attrition? The Effect of Recent L1 Re-exposure on Spanish Speakers under L1 Attrition. *Bilingual. Lang. Cognit.* 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. Phonetics* 40, 249–268. doi: 10.1016/j.wocn.2011.10.007

Chang, C. B. (2013). A Novelty Effect in Phonetic Drift of the Native Language. *J. Phonet.* 41, 520–533. doi: 10.1016/j.wocn.2013.09.006

Chang, Y. F. (2009). How to say no: an analysis of cross-cultural difference and pragmatic transfer. *Lang. Sci.* 31, 477–493. doi: 10.1016/j.langsci.2008.01.002

Cherciov, M. (2011). Between attrition and acquisition: The dynamics between two languages in adult migrants. *Dissert. Abstr. Int. Sect. A* 73:1761.

Chunpeng, C., and Hee-Don, A. (2017). The Influence of L2 on L1 Lexical Competence and Grammatical Properties in Written Expressions. *Chin. J. Appl. Linguist.* 40, 181–191. doi: 10.1515/cjal-2017-0011

Clahsen, H., and Felser, C. (2006). How native-like is non-native language processing? *Trends Cognit. Sci.* 10, 564–570. doi: 10.1016/j.tics.2006.10.002

Cohen, J. (1988). "Statistical Power Analysis for the Behavioral Sciences," in *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn, ed. J. Cohen (New Jersey, NJ: Lawrence Earlbaum Associates), doi: 10.4324/9780203771587

Contemori, C., Asiri, O., and Perea Irigoyen, E. D. P. (2019). Anaphora resolution in L2 English: an analysis of discourse complexity and cross-linguistic interference. *Stud. Second Lang. Acquisit.* 2019, 1–28. doi: 10.1017/S0272263119000111

Costa, A., and Sebastián-Gallés, N. (2015). How does the bilingual experience sculpt the brain? *Nat. Rev. Neurosci.* 15, 336–345. doi: 10.1038/nrn3709

Cunnings, I. (2017a). Parsing and working memory in bilingual sentence processing. *Bilingual. Lang. Cognit.* 20, 659–678. doi: 10.1017/S1366728916000675

Cunnings, I. (2017b). Interference in Native and Non-Native Sentence Processing. *Bilingual. Lang. Cognit.* 20, 712–721. doi: 10.1017/S1366728916001243

Daneman, M., and Green, I. (1986). Individual differences in comprehending and producing words in context. *J. Mem. Lang.* 25, 1–18. doi: 10.1016/0749-596x(86)90018-5

de Bot, K., Lowie, W., and Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingual. Lang. Cognit.* 10, 7–21. doi: 10.1017/S1366728906002732

De Cat, C., Gusnanto, A., and Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Stud. Sec. Lang. Acquisit.* 40, 119–151. doi: 10.1017/S0272263116000486

Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosop. Transact. R. Soc. B* 369:20120394. doi: 10.1098/rstb.2012.0394

Diependaele, K., Duñabeitia, J. A., Morris, J., and Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *J. Mem. Lang.* 64, 344–358. doi: 10.1016/j.jml.2011.01.003

Domínguez, L. (2013). *Understanding interfaces: Second language acquisition and first language attrition of Spanish subject realization and word order variation*. Amsterdam: John Benjamins.

Dragoy, O., Virfel, E., Yurchenko, A., and Bastiaanse, R. (2019). Aspect and tense attrition in Russian-German bilingual speakers. *Int. J. Bilingual.* 23, 275–295. doi: 10.1177/1367006917728388

Dussias, P. E. (2003). Syntactic ambiguity resolution in second language learners: Some effects of bilinguality on L1 and L2 processing strategies. *Stud. Sec. Lang. Acquisit.* 25, 529–557. doi: 10.1017/s0272263103000238

Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *Int. J. Bilingual.* 3, 355–371. doi: 10.1177/13670069040080031001

Dussias, P. E., and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingual. Lang. Cognit.* 10:101.

Dussias, P. E., Perrotti, L., Carlson, M., and Morales, L. (2016). "Exposure to a second language can change processing routines in the first language," in *Paper presented at the International Meeting of the Psychonomic Society, Granada, Spain. May 5–8*, (Granada: IMPS).

Duval, S., and Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–463. doi: 10.1111/j.0006-341x.2000.00455.x

Elston-Güttler, K. E., Gunter, T. C., and Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognit. Brain Res.* 25, 57–70. doi: 10.1016/j.cogbrainres.2005.04.007

Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., and Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Mem. Cognit.* 33, 1220–1234. doi: 10.3758/bf03193224

Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *J. Cognit. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377

Gunnerud, H. L., ten Braak, D., Reikerås, E. K. L., Donolato, E., and Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychol. Bull.* 146, 1059–1083. doi: 10.1037/bul0000301

Gürel, A. (2015). First language attrition of constraints on wh-scrambling: Does the second language have an effect?. *Int. J. Bilingual.* 19, 75–91. doi: 10.1177/1367006913506131

Hatzidaki, A., Branigan, H. P., and Pickering, M. J. (2011). Co-activation of syntax in bilingual language production. *Cognit. Psychol.* 62, 123–150. doi: 10.1016/j.cogpsych.2010.10.002

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *J. Educat. Statist.* 6, 107–128. doi: 10.3102/10769986006002107

Hedges, L. V., Tipton, E., and Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Res. Synth. Methods* 1, 39–65. doi: 10.1002/jrsm.5

Hicks, G., and Domínguez, L. (2020). A model for L1 grammatical attrition. *Sec. Lang. Res.* 36, 143–165. doi: 10.1177/0267658319862011

Higgins, J., and Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]*. Available online at: www.cochrane-handbook.org

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.

Kaan, E., Futch, C., Fernández Fuertes, R., Mujcinovic, S., Álvarez, and De La Fuente, E. (2019). Adaptation to syntactic structures in native and nonnative sentence comprehension. *Appl. Psycholinguist.* 40, 3–27. doi: 10.1017/S0142716418000437

Kaltsa, M., Tsimpli, I. M., and Rothman, J. (2015). Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua* 164, 266–288. doi: 10.1016/j.lingua.2015.06.002

Karayayla, T., and Schmid, M. S. (2019). First language attrition as a function of age at onset of bilingualism: First language attainment of Turkish–English Bilinguals in the United Kingdom. *Lang. Learn.* 69, 106–142. doi: 10.1111/lang.12316

Kasparian, K., and Steinhauer, K. (2017). When the second language takes the lead: Neurocognitive processing changes in the first language of adult attriters. *Front. Psychol.* 8:1–22. doi: 10.3389/fpsyg.2017.00389

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. *Cognit. Sci.* 41, 1760–1803. doi: 10.1111/cogs.12450

Kaushanskaya, M., and Smith, S. (2016). Do grammatical–gender distinctions learned in the second language influence native-language lexical processing? *Int. J. Biling.* 20, 30–39. doi: 10.1177/1367006915576830

Kremin, L. V., and Byers-Heinlein, K. (2021). Why not both? Rethinking categorical and continuous approaches to bilingualism. *Int. J. Bilingual.* 2021:13670069211031986. doi: 10.1177/13670069211031986

Kroll, J. F., Bobb, S. C., and Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Curr. Direct. Psychol. Sci.* 23, 159–163. doi: 10.1177/0963721414528511

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

Laufer, B., and Baladzhaeva, L. (2015). First language attrition without second language acquisition. *ITL* 166, 229–253. doi: 10.1075/itl.166.2.02lau

Levy, B. J., McVeigh, N. D., Marful, A., and Anderson, M. C. (2007). Inhibiting your native language: the role of retrieval-induced forgetting during second-language acquisition. *Psychol. Sci.* 18, 29–34. doi: 10.1111/j.1467-9280.2007.01844.x

Li, P., and Clariana, R. (2019). Reading comprehension in L1 and L2: An integrative approach. *J. Neurolinguist.* 50, 94–105. doi: 10.1016/j.jneuroling.2018.03.005

Libben, M. R., and Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *J. Exp. Psychol.* 35, 381–390. doi: 10.1037/a0014875

Liceras, J., and Senn, C. (2009). Linguistic theory and the analysis of minority languages: native, immigrant and heritage Spanish speakers. *Lengua Migración Lang. Migrat.* 1, 39–74.

Linck, J. A., Kroll, J. F., and Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning. *Psychol. Sci.* 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x

Luk, G., and Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *J. Cognit. Psychol.* 25, 605–621. doi: 10.1080/20445911.2013.795574

Macizo, P., Bajo, T., and Martín, M. C. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish–English interlexical homographs. *J. Mem. Lang.* 63, 232–244. doi: 10.1016/j.jml.2010.04.002

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Sec. Lang. Res.* 19, 144–161. doi: 10.1191/0267658303sr217ra

Misra, M., Guo, T., Bobb, S. C., and Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *J. Mem. Lang.* 67:1. doi: 10.1016/j.jml.2012.05.001

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

Monaghan, P., Chang, Y. N., Welbourne, S., and Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *J. Mem. Lang.* 93, 1–21. doi: 10.1016/j.jml.2016.08.003

Ooi, S. H., Goh, W. D., Sorace, A., and Bak, T. H. (2018). From Bilingualism to Bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingual. Lang. Cognit.* 21, 867–879. doi: 10.1017/S1366728918000020

Opitz, C. (2010). L1 attrition and L2 acquisition: Global language proficiency and language dominance in adult bilinguals. *EUROSLA Yearb.* 10, 248–281. doi: 10.1075/eurosla.10.13opi

Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2015). Rayyan — a web and mobile app for systematic reviews. *Systemat. Rev.* 5:210. doi: 10.1186/s13643-016-0384-4

Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., and Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *Quart. J. Exp. Psychol.* 63, 1631–1645. doi: 10.1080/17470210903511210

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Iss. Appl. Linguist.* 11, 175–205.

Pavlenko, A. (2004). "Second language influence and first language attrition in adult bilingualism," in *First language attrition: Interdisciplinary perspectives on methodological issues*, eds M. Schmid, B. Köpke, M. Kejser, and L. Weilemar (Amsterdam: John Benjamins), 47–59. doi: 10.1075/sibil.28.04pav

Pavlenko, A., and Malt, B. C. (2011). Kitchen Russian: Cross-linguistic differences and first-language object naming by Russian-English bilinguals. *Bilingual. Cognit.* 14, 19–45. doi: 10.1017/s136672891000026x

Perfetti, C., and Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Sci. Stud. Reading* 18, 22–37. doi: 10.1080/10888438.2013.827687

Peristeri, E., Tsimpli, I., Sorace, A., and Tsapkini, K. (2018). Language interference and inhibition in early and late successive bilingualism. *Bilingual. Lang. Cognit.* 21, 1009–1034. doi: 10.1017/S1366728917000372

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Ribbert, A., and Kuiken, F. (2010). L2-induced changes in the L1 of Germans living in the Netherlands. *Bilingual. Lang. Cognit.* 13, 41–48. doi: 10.1017/S1366728909990320

Ribes, Y., and Llanes, À (2015). First Language Attrition: The Effects of Acculturation to the Host Culture. *Proc. Soc. Behav. Sci.* 173, 181–185. doi: 10.1016/j.sbspro.2015.02.050

Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Sec. Lang. Res.* 28, 113–127. doi: 10.1177/0267658311418416

Román, P., Gonzáles, J., Ventura-Campos, N., Rodríguez-Pujadas, A., Sanjuán, A., and Ávila, C. (2015). Neural differences between monolinguals and early bilinguals in their native language during comprehension. *Brain Lang.* 150, 80–89. doi: 10.1016/j.bandl.2015.07.011

Román, P., Kaan, E., and Dussias, P. E. (2021). Access to verb bias and plausibility information in bilinguals. *Bilingual. Lang. Cognit.* 2021, 1–13. doi: 10.1017/S1366728921000924

Runnqvist, E., Gollan, T. H., Costa, A., and Ferreira, V. S. (2013). A disadvantage in bilingual sentence production modulated by syntactic frequency and similarity across languages. *Cognition* 129, 256–263. doi: 10.1016/j.cognition.2013.07.008

Sanoudaki, E., and Thierry, G. (2015). Language non-selective syntactic activation in early bilinguals: the role of verbal fluency. *Int. J. Bilingual Educat. Bilingual.* 18, 548–560. doi: 10.1080/13670050.2015.1027143

Schmid, M. (2010). Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingual. Lang. Cognit.* 13, 1–7. doi: 10.1017/S1366728909990368

Schmid, M. S. (2014). The debate on maturational constraints in bilingual development: A perspective from first-language attrition. *Lang. Acquisit.* 21, 386–410. doi: 10.1080/10489223.2014.892947

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 Karayayla, T. (2020). The Roles of Age, Attitude, and Use in First Language Development and Attrition of Turkish–English Bilinguals. *Lang. Learn.* 70, 54–84. doi: 10.1111/lang.12361

Schmid, M. S., and Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguist. Approac. Bilingual.* 7, 637–667. doi: 10.1075/lab.17058.sch

Schmid, M. S., Köpke, B., and de Bot, K. (2013). Language attrition as a complex, non-linear development. *Int. J. Bilingual.* 17, 675–682. doi: 10.1177/1367006912454619

Schmitz, K., Di Venanzio, 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. (2000). "Differential effects of attrition in the L1 syntax of near-native L2 speakers," in *BUCLD 24: Proceedings of the 24th annual Boston University Conference on Language Development*, eds S. C. Howell, S. A. Fish, and T. K. Lucas (Somerville, MA: Cascadilla Press), 719–725.

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguist. Approac. Bilingual.* 1, 1–33. doi: 10.1075/lab.1.1.01sor

Sorace, A. (2020). L1 attrition in a wider perspective. *Sec. Lang. Res.* 36, 203–206. doi: 10.1177/0267658319895571

Stasenko, A., and Gollan, T. H. (2019). Tip of the tongue after any language: Reintroducing the notion of blocked retrieval. *Cognition* 193:104027. doi: 10.1016/j.cognition.2019.104027

StataCorp (2015). *Stata Statistical Software: Release 14.2* (14.2). College Station, TX: StataCorp LP.

Steinhauer, K., and Kasparian, K. (2020). Brain Plasticity in Adulthood—ERP Evidence for L1-attrition in Lexicon and Morphosyntax After Predominant L2 Use. *Lang. Learn.* 70, 171–193. doi: 10.1111/lang.12391

Stuck, A. E., Rubenstein, L. Z., Wieland, D., Vandenbroucke, J. P., Irwig, L., Macaskill, P., et al. (1998). Bias in meta-analysis detected by a simple, graphical. *BMJ* 316, 629–634. doi: 10.1136/bmj.316.7129.469

Tanner-Smith, E. E., and Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Res. Synth. Methods* 5, 13–30. doi: 10.1002/jrsm.1091

Tanner-Smith, E. E., Tipton, E., and Polanin, J. R. (2016). Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R. *J. Dev. Life Course Criminol.* 2, 85–112. doi: 10.1007/s40865-016-0026-5

Titone, D., Libben, M., Mercier, J., Whitford, V., and Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *J. Exp. Psychol.* 37, 1412–1431. doi: 10.1037/a0024492

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

Walenski, M., Europa, E., Caplan, D., and Thompson, C. K. (2019). Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 40, 2275–2304. doi: 10.1002/hbm.24523

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., and Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159–170. doi: 10.1016/s0896-6273(02)01150-9

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

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 14 March 2022
doi: 10.3389/fpsyg.2022.716554

On the Reliability of the Notion of Native Signer and Its Risks

*Giorgia Zorzi[1], Beatrice Giustolisi[2], Valentina Aristodemo[3], Carlo Cecchetto[2,4], Charlotte Hauser[3], Josep Quer[1,5], Jordina Sánchez Amat[1] and Caterina Donati[3]**

[1] Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain, [2] Department of Psychology, University of Milano-Bicocca, Milan, Italy, [3] Université Paris Cité, CNRS, Laboratoire Linguistique Formelle UMR 7110, Paris, France, [4] UMR 7023 Structures Formelles du Langage, Paris, France, [5] ICREA – Pompeu Fabra University, Barcelona, Spain

OPEN ACCESS

Edited by:
Cristina Suarez-Gomez, University of the Balearic Islands, Spain

Reviewed by:
Wolfgang Mann, University of Roehampton London, United Kingdom
Isabel Reyes Rodríguez-Ortiz, Seville University, Spain

***Correspondence:**
Caterina Donati
caterina.donati@u-paris.fr

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

Received: *28 May 2021*
Accepted: *10 January 2022*
Published: *14 March 2022*

Citation:
Zorzi G, Giustolisi B, Aristodemo V, Cecchetto C, Hauser C, Quer J, Sánchez Amat J and Donati C (2022) On the Reliability of the Notion of Native Signer and Its Risks. Front. Psychol. 13:716554. doi: 10.3389/fpsyg.2022.716554

Who is a native signer? Since around 95% of deaf infants are born into a hearing family, deaf signers are exposed to a sign language at various moments of their life, and not only from birth. Moreover, the linguistic input they are exposed to is not always a fully fledged natural sign language. In this situation, is the notion of native signer as someone exposed to language from birth of any use? We review the results of the first large-scale cross-linguistic investigation on the effects of age of exposure to sign language. This research involved about 45 Deaf adult signers in each of three sign languages (Catalan Sign Language, French Sign Language, and Italian Sign Language). Across the three languages, participants were divided into three groups – those exposed from birth, those between 1 and 5 years of age, and those exposed between 6 and 15 years of age – and received a battery of tests designed for each language targeting various aspects of morphosyntactic competence. In particular, the tests focused on both those morphosyntactic phenomena that are known from the spoken language literature to be good detectors of language impairment or delay (i.e., *wh*-interrogatives and relative clauses) and on morphosyntactic phenomena that are sign language specific (i.e., role shift and directional verbs). The results showed a clear effect of being native, with significant differences across languages and tests between signers exposed to sign language from birth and those exposed in the 1st years of life. This confirms the life-long importance of language exposure from birth and the reliability of the notion of "*nativeness*", at least for syntax. On the other hand, while in most domains the differences observed between populations might be differences in performance, for some specific constructions, signers belonging to the three groups may have different grammars. This latter finding challenges the generalized use of native signers' grammar as the baseline for language description and language assessment.

Keywords: sign languages, native signer, early and late signers, effect of age of exposure, language assessment

INTRODUCTION

The notion of "native" user of a language has become controversial for various reasons. For spoken languages, the challenge comes from bilingualism and multilingualism (Sorace, 2021 for an overview), while for sign languages the controversy is due to the unique sociolinguistic situation that characterizes the population of Deaf signers.[1] The linguistic profiles attested among deaf people are very diverse, and native signers, defined as deaf individuals who were born into a Deaf signing family, are only a small minority. This led many scholars to challenge the importance of this notion as a reliable criterion for language description and assessment, at least as far as sign languages are concerned. The question is whether "*nativeness*" is indeed different from early exposure: in other words, whether what really matters is being early exposed to a sign language, or whether there is a special status associated to being native, even with respect to early learners.

In this paper we will first discuss the controversial status of "native signers" with respect to the global population of Deaf signers, underlying that most experimental studies have not been using consistent criteria to contrast native signers from those signers who were not exposed to a sign language from birth. With the goal of contributing with experimental methods to the debate of whether native and non-native signers are indeed different, in the section "Native, Early and Late Signers in a Large-Scale Cross-Linguistic Investigation" we will present an overview of the morpho-syntactic comprehension tests developed within the Horizon 2020 project SIGN-HUB ("The Sign Hub: preserving, researching and fostering the linguistic, historical and cultural heritage of European Deaf signing communities with an integral resource") describing the criteria that were used to select the groups of native, early and late signers, and the tests themselves, and providing a summary of the results. In the section "Discussion", we will then discuss the results presented in the previous section and support the claim that native signers have indeed a different performance in comparison to signers who were exposed later to sign language, even early in life. At the same time, we will challenge the reliability of native signers' grammar as the baseline to be used for sign language investigation and assessment.

THE CONTROVERSIAL NOTION OF "NATIVE SIGNER"

The population of deaf pre-lingual adult signers is extremely heterogeneous, as it is characterized by individuals with very different linguistic backgrounds (Johnston, 2006; Costello et al., 2008; Quer and Steinbach, 2019). This is due to the sociolinguistic situation that characterizes deaf people and Deaf communities. The general estimation is that only 5–10% of deaf babies have deaf signing parents, and even less have deaf signing grandparents (Newport, 1988; Neidle et al., 2000; Mitchell and Karchmer, 2004) and therefore only a small part of the deaf population is exposed to sign language from birth. This percentage, however small, has been calculated on the deaf population of the United States, but it has been questioned as an overestimation for the deaf population in Australia (Johnston, 2006) and Europe (Costello et al., 2008). Costello et al. (2008) looked at deaf signers in the Basque Country, in the north of Spain, underlining that the number of deaf people born into Deaf families is extremely low and it hardly reaches 5%.[2] This aspect needs to be taken into consideration especially when looking at deaf populations in smaller communities.

How Deaf Children Get Exposed to Language

If we consider the general definition of native signers as "Deaf people who grew up with Deaf signing parents and who identify with the Deaf community" (Neidle et al., 2000), it is clear that it refers to a very small part of the deaf population. In addition to this, it is important to remark that some deaf parents might have been themselves exposed to sign language at a late point in life and therefore might provide a language input to the child that cannot be strictly compared to the one of a native (Lillo-Martin, 2021). Even if it has been shown that deaf children exposed to a non-native sign language from birth reach a better performance than their parents and get close to their native peers, they are still not native-like (Singleton and Newport, 2004). Deaf children exposed to a native input might thus be exceedingly rare.

As for the rest of the deaf children population, it is constituted of deaf children born in hearing families, and therefore for the most part they are not exposed to sign language from birth. There are several factors that prevent the deaf population from being exposed to an early and adequate sign language input that would allow an early and natural language acquisition. The main reasons are: the age of diagnosis, although it has recently drastically decreased due to newborn hearing screening (Joint Committee on Infant Hearing, 2019), the different degrees of deafness, and the use of technologies such as cochlear implants or hearing aids, together with the type of language intervention adopted by the parents: they might opt for exposure to spoken language alone *via* amplification through hearing aids or cochlear implants, or rather for exposure to both spoken and sign languages or for exposure to sign language only. In many cases parents are advised by doctors and educators to adopt an oralist approach supporting the use of technologies meant to facilitate the learning of spoken languages, denying sign input (Lillo-Martin, 2021). Even with early intervention through hearing aids or cochlear implants, though, language access is delayed if not provided through a fully accessible input, which in the case of deaf children is in the visual-gestural modality (Humphries et al., 2016; Hall et al., 2019, among others). In a very small percentage of cases, hearing parents decide to learn sign language and expose their child to it (cf. Chen Pichler and Lillo-Martin, 2018), hence still delaying giving a sign language input while they go through the process of learning the

[1] Following standard practice, in this paper we use the word "deaf" with low case "d" to refer to the audiological condition of a deaf individual. The word "Deaf" with capital "D", instead, will be used to refer to the deaf members of a community who use a sign language, with its own cultural identity.

[2] Costello et al. (2008) used data provided by the Spanish National Association of the Deaf and the Spanish Statistical Institute.

language. Eventually, the input they provide cannot be compared to the one of deaf native signers (Lillo-Martin, 2021), even though it is provided early in life. Only a very small minority of deaf children born in hearing families is then exposed to sign language in their parental home shortly after diagnosis. In most cases, it is only in school that deaf children get exposed to sign language.

A delayed exposure to sign language leads to a delay in the development of language, and even to atypical neurological mappings of language (Mayberry, 2010; Mayberry and Kluender, 2018; Woll, 2018). Moreover, it has sociolinguistic consequences in relation to how deaf people not exposed to sign language from birth relate to the Deaf community. For German Sign Language (DGS), Jaeger (2019) distinguishes between a "native" and an "authentic" signer. Many participants to her study who were non-native reported that they identify themselves as "authentic signers", specifying that such status can be reached either by being born into the Deaf community ("Deaf aristocracy"), or *via* intentional change ("Deaf meritocracy"), as was the case for most of them. This perspective on non-native signers as being native-like from an identity perspective relates to the conceptualization of non-native signers as "New Signers", a concept adapted from that of "New Speaker". The term "New Speaker" was introduced to indicate users who acquire a minority language later in life and outside the parental home (O'Rourke et al., 2015), especially in the context of language revitalization (Jaffe, 2015). It has been recently extended to deaf non-native signers since they share the characteristics of acquiring language after childhood and outside the parental home (Jaeger, 2019) and because of the status of sign languages as minority languages (Bauman and Murray, 2017; Tupi, 2019). The New Signer model gives a new perspective to the "native speaker" ideology and shows that it is important to disentangle sociological and psycholinguistic factors when it comes to identifying the profile of a native signer.

Studying a sign language by only relying on native signers might end up as an impossible task. The alternative that has been adopted in the literature is to work with consultants that fulfill several criteria that make them as close as possible to the standard definition of native signers (Quer and Steinbach, 2019). As reported by Costello et al. (2008), many research groups tend to select participants, especially for neurolinguistic studies, who are (at least) second generation deaf-of-deaf signers. On the other hand, Mathur and Rathmann (2006) consider three main criteria: (i) exposure to sign language by the age of three; (ii) ability to give grammaticality judgments with ease; and (iii) daily contact with a sign language in the Deaf community for more than 10 years. In experimental data assessing language acquisition and the impact of age of exposure (AoE) on language competence, native signers tend to be strictly identified with individuals who have been exposed to sign language from birth from Deaf signing parents. Oftentimes, though, a limit of 3 years of age is established to consider someone as native (Mayberry, 1993; Freel et al., 2011).

It is clear that determining the exact criteria that define an individual as having native competence is particularly crucial when the aim is to assess the consequences that a delay to language exposure can cause earlier or later in life. In the following section, we provide an overview of the profiles of deaf signers that have been studied in this type of studies. In many cases, their goal is to determine whether native signers, even if they constitute a minority, can be distinguished from signers who have been exposed to sign language even quite early in life, as far as language development is concerned.

Age of Exposure to Sign Language

Early exposure to language is crucial for language acquisition (Mayberry et al., 2002) and this has been documented for sign languages since the '90s, with studies showing that non-native signers differ from native signers in several morpho-syntactic tasks. Emmorey et al. (1995), in a study on sign recognition within a sentence containing errors in verb agreement, showed that only native signers were sensitive to agreement errors, while late learners were not. The relevant group of late learners were exposed to American Sign Language (ASL) between 4 and 20 years of age. In a second experiment involving sign recognition in a sentence containing errors in verb agreement or aspect and offline grammaticality judgments, non-native signers were distinguished into early and late learners, with AoE range of 2–7 years and 10–20 years, respectively. The results of the first experiment were confirmed, with native signers outperforming non-native regardless of their AoE group. In other studies on ASL, though, the AoE effect was gradient, showing a continuum across the groups: as AoE increased, the performance of signers decreased. This is the case of a study on ASL sentence processing measured by recall of long and complex sentences. In this study, Mayberry (1993) included three groups of pre-lingual deaf signers with AoE ranging from (i) 0–3 years of age, (ii) 5–8 years, and (iii) 9–13 years (and a fourth group of post-lingual deaf signers who were exposed to ASL between 8 and 15 years of age and lost their hearing between 8 and 12 years of age). The performance of the pre-lingual deaf signers decreased as AoE increased. A similar result was obtained using a grammaticality judgment task on sentences of various types, independently from the syntactic structure investigated (Boudreault and Mayberry, 2006). In the same task, reproduced by Cormier et al. (2012) in British Sign Language (BSL), accuracy in the grammaticality judgment task decreased as AoE increased for Deaf early signers, while no decreasing related to AoE was found among Deaf late signers. However, if we compare the AoE of late learners in the two versions of the study, we observe that while in the ASL experiment late learners were exposed to ASL between 8 and 13 years, in the BSL experiment late learners were exposed to BSL between 9 and 18 years. More importantly, late ASL signers were described as L1 signers, whereas Cormier et al. (2012) suggest that their group of late signers was composed of L2 signers, with English as L1. The upward trend for the oldest AoE was then attributed to having acquired another language from birth.

The characteristics of the various groups of signers participating in the experiments just presented are summarized in **Table 1**.

From **Table 1**, focusing on AoE, we can clearly see that there is a lot of variation across studies on the groups of signers investigated and how they are defined: in some cases, native signers are compared directly to late learners. In other cases, when three populations are indeed distinguished, the AoE range of the three groups varies a lot. It is possible to see variation in

TABLE 1 | Summary of the age of exposure (AoE) of participants across a selection of relevant studies on the impact of AoE.

Language	Task	Participants
ASL	Sign recognition in a sentence containing errors in verb agreement (Emmorey et al., 1995)	(i) 11 Native: AoE = birth, Age = 21–44 (ii) 10 Late: AoE = 4–20 (*M* = 12), Age = 29–49
	Sign recognition in a sentence containing errors in verb agreement or aspect and offline grammaticality judgments (Emmorey et al., 1995)	(i) 10 Native: AoE = birth, Age = 19–24 (ii) 10 Early: AoE = 2–7 (*M* = 4), Age = 21–37 (iii) 10 Late: AoE = 10–20 (*M* = 14), Age = 22–46
	Sentence processing (Mayberry, 1993)	(i) 9 AoE = 0–3 (*M* = birth), M Age = 51 (43–67), M SLe = 51 (43–67), born deaf (ii) 9 AoE = 5–8 (*M* = 7), M Age = 61 (37–71), M SLe = 51 (31–65), born deaf (iii) 9 AoE = 9–13 (*M* = 11), M Age = 60 (40–72), M SLe = 54 (28–61), born deaf (iv) 9 AoE = 8–15 (*M* = 11), M Age = 60 (38–72), M SLe = 50 (29–61), onset deafness: 8–12 (*M* = 9)
	Grammaticality judgment task on sentences (Boudreault and Mayberry, 2006)	(i) 10 Native: AoE = birth, M age = 24.2 (18–41), M SLe = 24.3 (18–41) (ii) 10 Early: AoE = 5–7 (*M* = 5.6), M Age = 43.2 (31–62), M SLe = 37.6 (14–47) (iii) 10 Late: AoE = 8–13 (*M* = 10.3), M Age = 43 (24–79), M SLe = 32.9 (13–71)
BSL	BSL version of Boudreault and Mayberry's task (Cormier et al., 2012)	(i) 10 Native: AoE = birth, M Age = 39.7 (20–57), M SLe = 39.7 (20–57) (ii) 11 Early: AoE = 2–8 (*M* = 4.4), M Age = 36.5 (19–54), M SLe = 32 (17–51) (iii) 9 Late: AoE = 9–18 (*M* = 12.8), M Age = 30.9 (20–43), M SLe = 18.1 (10–26)

the definition of early and late signers even in the "replication" of the same study (cf. Boudreault and Mayberry, 2006, and Cormier et al., 2012). Moreover, in Mayberry (1993) the category with the earliest exposure to ASL includes signers who were exposed before 3 years of age, without excluding native signers from this sample.

Under these circumstances, it is thus not straightforward to compare the results of the various studies. In particular, it is not clear whether the effect of AoE found in the literature so far is a simple effect of being early exposed to a sign language, or whether there is a special status associated to being exposed from birth even with respect to early learners.

NATIVE, EARLY AND LATE SIGNERS IN A LARGE-SCALE CROSS-LINGUISTIC INVESTIGATION

A large-scale cross-linguistic investigation was conducted within the SIGN-HUB project. With the aim of investigating the role of AoE in language comprehension in adulthood, four morpho-syntactic comprehension tests were developed in three different sign languages (Catalan Sign Language (LSC), French Sign Language (LSF), and Italian Sign Language (LIS)]. Results of those tests, separately discussed in Aristodemo et al. (2020, in press), Cecchetto et al. (2021), Hauser et al. (2021, in press), are crucial to understanding whether native and non-native signers differ categorically, or whether what matters is simply early exposure to sign language for which we expect a gradient effect associated to different AoE groups.

The Participants in the SIGN-HUB Tests

In the SIGN-HUB tests, for the three languages (LIS, LSC, and LSF), participants were selected following three general inclusion criteria: (i) onset of deafness not later than 3 years of age;[3] (ii) first exposure to sign language not later than 15 years of age; and (iii) the target sign language as their preferred mean of communication. All participants had been exposed to sign language for at least 15 years, with the exception of two young LSF participants, who both had only 9 years of sign language experience.

To be able to create groups of participants with a similar language input and background, they were asked to fill in a questionnaire containing several personal questions including AoE, the possible deafness of their parents, whether their parents were signers, whether they went to a school for the deaf or had deaf school mates, and so on.[4] Participants were divided into three groups: (i) native, (ii) early, and (iii) late signers. Native signers were individuals exposed to sign language from birth (AoE = 0), having at least one deaf signing parent, and who therefore acquired SL in a family environment. Early learners were exposed to sign language between 1 and 5 years of age while late learners between 6 and 15 years of age. The choice of the age ranges was based (i) on including among native signers only those people exposed to a sign language from birth; (ii) on having in the early learners group signers who were exposed to a sign language very early in life or at least within the critical acquisition period up to 5 years of age, but not from birth; and (iii) on comprising in the late learners group signers who were exposed to sign language not later than 15 year old, which is the average age limit for being exposed to a sign language in a school setting in the target language countries. In both groups of non-native signers, most participants were introduced to sign language in institutional educational settings (preschool for early signers and school for late signers), almost none had deaf parents, and very few had at least one parent knowing sign language. **Table 2** summarizes the

[3]Concerning the onset of deafness, participants self-reported that it was never later than 3 years old (LIS: *M* = 3.5 months, LSC: *M* = 5.6 months, LSF: *M* = 3.7 months).

[4]Questionnaires were written, but a signing person was present so participants who had doubts could ask for a translation.

TABLE 2 | Summary of SIGN-HUB participants' characteristics per group and language.

Group	SL	N.	AoE	Everyday use of SL	Deaf parent(s)	Signing parent(s)	Context of exposure to SL	Years of SL experience
NATIVE	LIS	16	0	16	16	16	Family: 16	30–60 (*M* = 43)
	LSC	14	0	13[a]	14	14	Family: 14	26–69 (*M* = 44)
	LSF	14	0	13[b]	13	13	Family: 13 (1 NS)	26–54 (*M* = 39)
EARLY	LIS	15	2–5 yrs (*M* = 3,9)	13	1	3	Family: 4 Preschool: 10 (1 NS)	32–58 (*M* = 47)
	LSC	16	3–5 yrs (*M* = 3.5)	15	1	2	Family: 3 Preschool: 13	20–60 (*M* = 48)
	LSF	15	1–5.5 yrs (*M* = 3.4)	10	none	1	Family: 3 Preschool: 11 (1 NS)	20–39 (*M* = 30)
LATE	LIS	13	6–15 yrs (*M* = 9.1)	11	none	1	Family: 2 School: 9 (2 NS)	26–58 (*M* = 41)
	LSC	12	6–15 yrs (*M* = 10.4)	11	1	2	School: 8 (4 NS)	34–57 (*M* = 41)
	LSF	14	6–14 yrs (*M* = 9.2)	11	2	1	Family: 1 School: 9 (4 NS)	9–63 (*M* = 31)

[a]*For LSC, one native signer, one early and one late declared to use LSC "often" instead of "everyday".*
[b]*For LSF, one native, five early and three late signers declared to use LSF "often" instead of "everyday".*

TABLE 3 | Summary of SIGN-HUB participants' general characteristics per group and language.

Group	SL	N.	Age	Gender	Degree of deafness[9]	Hearing aids	Education
NATIVE	LIS	16	30–60 (*M* = 43)	10 female 6 male	15 very severe 1 moderate	6 hearing aids	Median = high school
	LSC	14	26–69 (*M* = 44)	7 female 7 male	13 very severe 1 moderate	None	Median = university education
	LSF	14	26–54 (*M* = 39)	6 female 8 male	9 very severe 5 severe	7 hearing aids 1 cochlear implant	Median = middle school
EARLY	LIS	15	34–62 (*M* = 48)	7 female 9 male	14 very severe 1 severe	5 hearing aids	Median = high school
	LSC	16	23–64 (*M* = 51)	10 female 6 male	16 very severe	None	Median = middle school
	LSF	15	24–47 (*M* = 34)	10 female 5 male	13 very severe 2 severe	4 hearing aids 1 cochlear implant	Median = university education
LATE	LIS	13	40–65 (*M* = 50)	4 female 9 male	10 very severe 2 severe 1 moderate	3 hearing aids 1 cochlear implant	Median = high school
	LSC	12	41–63 (*M* = 52)	5 female 7 male	10 very severe 2 severe	5 hearing aids	Median = middle school
	LSF	14	19–72 (*M* = 40)	8 female 6 male	12 very severe 2 severe	6 hearing aids 1 cochlear implant	Median = high school

[9]*Following the recommendation by the International Bureau for Audiophonology BIAP, "very severe" is considered a degree of deafness higher than 90 dB, "severe" a degree between 71 and 90 dB, and "moderate" a degree of deafness between 41 and 70 dB.*

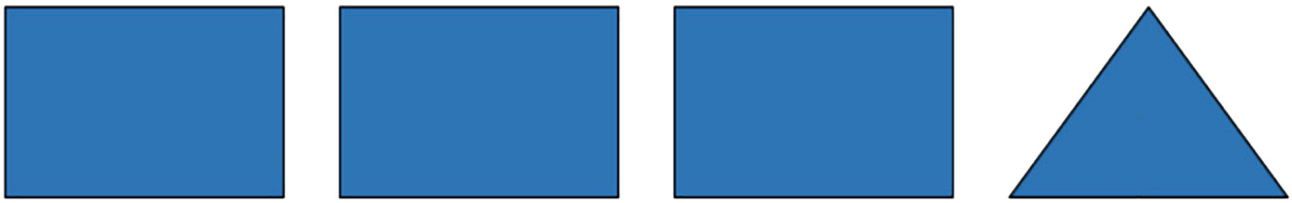

FIGURE 1 | Example of one item of the Odd One Out Cognitive Task (Hauser et al., 2021: 18). (CC-BY 4.0).

characteristics of all participants that were considered for the first participant selection.

The questionnaires participants filled in were also used to collect more general personal information. **Table 3** summarizes data of the final pool of participants considered in the analyses about chronological age (inserted as a factor in the various analyses), gender, degree of deafness, and use of hearing aids. The questionnaire was meant also to collect information about participants' use of written language (either Italian, Catalan, Spanish, or French). They were asked to self-rate whether they used written language every day, and if they read newspapers, etc. However, the data obtained, which might be considered as an indirect measure of their competence in the spoken language, were often not coherent and in any case not fine-grained enough to be used as a factor in the analyses.

All participants were screened for cognitive deficits, using the Odd One Out Cognitive Task (Giustolisi and Friedmann, 2019, for LIS, Zorzi et al., 2019c, for LSC and Aristodemo and Friedmann, 2019, for LSF), which was designed to detect potential cases of cognitive impairment. In this test, participants needed to find the intruder in a set of four pictures (see **Figure 1** for an example). The Odd One Out Cognitive Task displayed 28 items preceded by two training items. For each participant, *z*-scores were calculated considering language group mean and standard deviations. Participants with *z*-scores lower than −2.5 were excluded from the study.

One native participant was excluded both from the LIS pool and from the LSC pool. The LIS participant had a z-score of −3.94 and the LSC one a z-score of −5.51. No participant was excluded from the LSF pool.

The SIGN-HUB Tests

The SIGN-HUB project tests had two main goals: (i) providing data for the understanding of the effect of AoE in signers, and (ii) contribute to the comparative analysis of some specific linguistic phenomena. They were developed to study complex structures of two types: either characterized by long-distance dependencies and known to be good detectors of language impairment or delay (i.e., relative clauses and *wh*-questions), as in Friedmann et al. (2009), or prototypical sign language modality specific constructions (i.e., role shift and expression of agreement through directional verbs).[5] Each test was language specific, but they were similar in design and, most importantly, the criteria to distinguish the populations investigated were the same.[6]

Long Distance Dependencies: Relative Clauses and *Wh*-Questions

For head initial languages such as English it has been found that subject relative clauses are easier to understand than object relative clauses, and this is also the case for subject *wh*-interrogatives with respect to object *wh*-clauses (Friedmann et al., 2009, among others). Such asymmetry, that goes under the name of Subject Advantage, has been accounted in various ways, with proposals pointing at resource-based effects related to structural distance (Frazier, 1987; Hawkins, 1999), intervention (Friedmann et al., 2009), linear distance (e.g., King and Just, 1991; Gibson, 2000), canonical order effects (Diessel and Tomasello, 2005), distribution-based effects (e.g., Mak et al., 2006), and prominence-factors (Van Valin and Wilkins, 1996). Most studies point toward a universal Subject Advantage at the cross-linguistic level, but interestingly, most of them focus on head initial languages. In the SIGN-HUB project tests, LSF allows both SOV and SVO orders with preference varying across individuals (Hauser, 2019), while LIS and LSC show an SOV order (Quer, 2002; Cecchetto et al., 2006). Moreover, among the three languages, different strategies are used to realize relative clauses and *wh*-constructions: LSF has head-external relative clauses and *in situ wh*-interrogatives (Hauser, 2019), while LIS and LSC have head-internal relative clauses and *wh*-clauses involving *wh*-movement to the right periphery of the clause (Quer et al., 2005; Branchini and Donati, 2009; Cecchetto et al., 2009; Mosella Sanz, 2012). In addition to providing new results contributing to the debate of age of language exposure as a factor in language assessment, which we shall discuss here, the SIGN-HUB tests also provide crucial data from a different modality on how to explain the Subject Advantage from a theoretical point of view. We refer to Cecchetto et al. (2021), Hauser et al. (2021, in press) for a detailed discussion of these conclusions.

Concerning the SIGN-HUB tests on relative clauses (Hauser et al., 2021), they aimed at investigating the comprehension of subject and object relative clauses in a sentence-to-picture matching task based on Friedmann et al. (2009). In

[5]A secondary goal was to start developing clinical tests to assess language impairment in Deaf adults. With this purpose in mind, two lexical comprehension tasks were also designed in order to detect potential impairments at the lexical level, in the phonological system and in the semantic one, respectively. The question of the role of AoE was marginal in these tasks, since the literature suggests that AoE does not have an impact on the size of the lexicon (i.e., on accuracy in lexical comprehension tasks, e.g., Carreiras et al., 2008; Dye and Shih, 2009).

[6]All SIGN-HUB tests are available under request. For more information, see https://www.sign-hub.eu/assessment/welcome-page-assessment

each picture, three characters were displayed: two identical characters either performing an action or undergoing that action with respect to a third different character standing between them (**Figure 2**).[7]

The same picture was used to match a subject RC (i.e., "Choose the lion that licks the dog") or an object RC (i.e., "Choose the lion that the dog licks"). Participants were asked to choose one of the characters depending on the type of relative clause they were watching.

As for *wh*-interrogatives (also called content questions), the SIGN-HUB tests aimed at assessing comprehension of subject and object questions in a similar sentence-to-picture matching task (Cecchetto et al., 2021; Hauser et al., in press). The pictures also displayed three characters, like the one in **Figure 2**, and the answer to the question always targeted the characters on the sides in order to test the subject and object conditions (i.e., "Who licks the cow?" and "Who does the cow lick?," respectively).

Modality Specific Phenomena: Role Shift and Agreement

Two comprehension tests within the SIGN-HUB project were created to investigate two constructions that are modality specific: role shift and the expression of spatial agreement through directional verbs. Role shift is commonly used in sign languages and is particularly interesting for its semantic properties; spatial agreement consists in a strategy expressing agreement through articulation in space of the trajectory associated with the verb. This latter phenomenon has been studied in other sign languages showing an important impact of AoE (Emmorey et al., 1995; Cormier et al., 2012, among others). The two tests were language specific but had a similar design across languages (Aristodemo et al., 2020, in press).

Role shift (RS) is a construction commonly used in sign languages to report utterances or thoughts from the perspective of an agent distinct from the utterance speaker (Quer, 2011). It is signaled by specific non-manual markers that can slightly vary across languages, but that in general are characterized by body/head movement toward the locus in space assigned to the referent whose utterance or thought has been reported, and eye-gaze contact break with the actual addressee. Interestingly, when introduced by a verb like SAY, but also when no introducing predicate is used, role shift displays indexical shift: indexical expressions like the first-person pronoun (IX_1) retrieve their reference from the reported context. One of the main goals of these tests was to assess the comprehension of pairs of sentences with and without role shift with a first-person pronoun embedded under SAY. Like the other tests presented so far, this study on role shift was also meant to make a contribution to the debate on the theoretical nature of this structure. The tests were a sentence-to-picture matching task, in which participants were asked to pick one of two pictures matching a target sentence. Crucially, the choice depended on whether participants shifted (or not) the referent of the first-person pronoun in the target sentence.

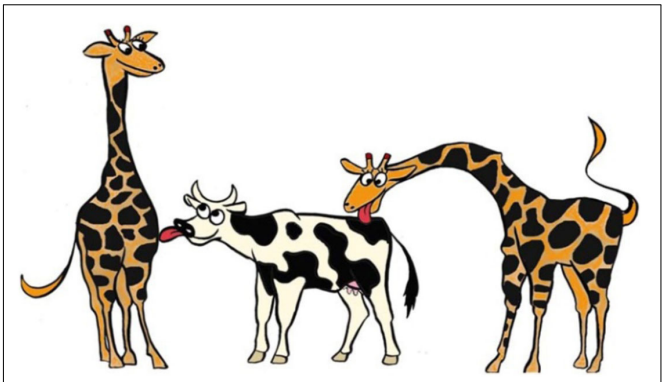

FIGURE 2 | Example of a three characters picture (Hauser et al., 2021: 14). (CC-BY 4.0).

As for the assessment of comprehension of agreement using directional verbs, this type of verbs are characterized by the articulation of a trajectory in the signing space from the locus associated with an argument toward the position associated with another argument. The SIGN-HUB tests were developed with the goal of assessing the comprehension of this phenomenon with a truth-value judgment task, in which the target sentence containing an agreeing verb appeared on the screen right after a non-linguistic clip describing a situation with at least two characters. Participants had to judge whether the target sentence matched the situation described in the clip or not (Aristodemo et al., 2020).

Summary of Results: Long Distance vs. Sign Language Specific

In all the tests the results were clear: a delayed AoE had a lifelong impact on individuals' language performance and/or competence.

As for the comprehension of *wh*-questions, only the results concerning LIS and LSF were analyzed so far. In LIS, native signers outperformed non-native not only in object questions, that were expected to be complex, but also in control questions, which were easy (Cecchetto et al., 2021). Even in this simple task, a difference emerged, confirming permanent effects of delayed exposure to sign language. For LSF, comparing language groups, a marginal difference was found between native and late learners, but a significant interaction emerged between this factor, the type of question and the subject/object condition. It was also found that the complexity provoked by object questions especially in which-questions particularly affected late learners of LSF. Importantly, in both LIS and LSF, early and late signers did not perform differently (Cecchetto et al., 2021; Hauser et al., in press).

The same consistent results have been found in the comprehension of role shift (RS) across the three languages: in LIS, native signers outperformed early and late signers

[7]Three characters pictures were used in the tests on relative clauses (Giustolisi et al., 2019, for LIS, Zorzi et al., 2019a, for LSC and Hauser et al., 2019, for LSF) and wh-interrogatives (Checchetto et al., 2019, for LIS, Zorzi et al., 2019b, for LSC and Aristodemo et al., 2019, for LSF).

both when the first-person pronoun appeared in subject and object position, with RS and without RS. In LSF, native signers outperformed early signers in all types of sentences in both conditions. Moreover, native signers outperformed late signers in all sentences with RS. This was not the case for sentences without RS, but one might speculate that this is because the late signers who performed worse in RS preferred by default the condition without RS. This might explain why late signers outperformed early signers in sentences without RS, and why the difference between late and native signers was not significant. In LSC, all groups had a good performance in sentences without RS. On the contrary, the performance in sentences with RS was more variable, but poor in native signers, and very poor in early and late learners. These results are attributed in the paper to a series of factors related to the non-manual markers associated to RS in the LSC tests, which were relatively subtle and might not have been clearly perceived by non-native signers (Aristodemo et al., in press).

The test on the comprehension of agreement with directional verbs, for which only LSF and LIS data were analyzed so far, also goes in the same direction: native signers outperformed non-native in LIS in the mismatch conditions. In LSF, instead, native signers were more accurate than non-native in both mismatch and match conditions. In general, no difference between early and late signers was found.

Finally, the test on relative clauses provides further evidence about the impact of AoE, and the special status of native signers. As for LSF, Hauser et al. (2021) report that for all three groups the difference between subject RCs (SRC) and object RCs (ORC) was significant, such that subject RCs were understood more easily. In the comprehension of ORCs, native signers performed significantly better than late learners and performed better than early learners in SRCs, but not significantly so. No significant difference was found between early and late learners. In LIS, native signers significantly outperformed early learners in SRCs, and they outperformed both early and late learners in ORCs. The difference between early and late learners was not significant.

As for LSC, the results obtained went even beyond expectations about AoE affecting adults' performance, and raised interesting questions. Again, SRCs were significantly better understood than object RCs across all three groups. As for ORCs, the difference between late and early learners only approached significance, while there was no significant difference between native signers and early learners. Late learners had a significantly lower performance than native signers. Interestingly, non-native learners were *below chance* when it came to ORCs, suggesting that non-native signers interpreted ORCs as SRCs. As discussed in detail in Hauser et al. (2021), this seems to represent an extreme case of AoE effect, where the difference in AoE produces a difference in grammar, not just in performance, with native signers having both SRCs and ORCs in their grammar while non-native signers not allowing ORCs at all in LSC.

The results we just outlined can be summarized in **Table 4**. For each language, the first column in **Table 4** indicates for every phenomenon investigated whether we found a significant difference in at least one condition of the tests between native and non-native signers. The second column summarizes for each phenomenon whether we found a significant difference in at least one condition of the tests between early and late learners.

TABLE 4 | Summary of the SIGN-HUB tests where native signers significantly outperformed non-native and where early learners significantly outperformed late learners in at least one condition of the tests.

	Native vs. Non-native			Early vs. Late		
	LIS	**LSC**	**LSF**	**LIS**	**LSC**	**LSF**
Wh-question comprehension	✓	NA	×	×	×	×
Role shift comprehension	✓	✓	✓	×	×	×
Directional verb comprehension	✓	NA	✓	×	×	×
Relative clause comprehension	✓	✓	✓	×	×	×

Table 4 clearly indicates that language exposure from birth is an important factor in determining language competence in the syntactic phenomena that were investigated. It also points at the importance of *nativeness* over simple earliness of first language exposure. These results have been obtained in sign languages that differ significantly in the syntactic domains under investigation. Nevertheless they are fully comparable as far as the effect of *nativeness* and AoE is concerned, since they have been obtained with comparable populations of signers divided according to the same criteria in three groups: native signers, defined as signers exposed to sign language from birth and with at least one Deaf signing parent; early learners, defined as been exposed to sign language between the age of 1 and 5 (included); late learners, defined as been exposed to sign language between the age of 6 and 15 years.

It is important to underline, though, that the effect of *nativeness* is not found to the same extent in every condition and in every language tested. This is mainly due to the amount of population tested (less than 15 people for each AoE group in each language). Nevertheless, the effect is overall consistent.

DISCUSSION

Summarizing the main findings of the tests described in the preceding section, we can conclude that a delayed AoE has a direct impact on syntactic competences. This conclusion holds both for those linguistic phenomena that are widely known to be sensitive to language acquisition disruption, such as the comprehension of long-distance dependencies (assessed in the SIGN-HUB tests through the comprehension of relative clauses and content questions), and for grammatical features that are more specific to the signing modality, such as the comprehension of role shift and of agreement with directional verbs. Moreover, this conclusion holds true across different sign languages, notwithstanding important syntactic differences across constructions.

Remember that the question at stake in this paper is whether the traditional centrality that is assigned to native

signers in the linguistic literature makes sense in relation to the signing populations, where native signers are a small minority, certainly not representative of the general population of signers.

Native Signers Are Different

In all the phenomena that were investigated, a significant difference emerged between native and early learners. This pattern appears to strongly confirm that there is a categorial effect of being a native signer that goes beyond simple AoE, a more continuous measure. There are at least two possible interpretations for this finding, not necessarily mutually exclusive.

A first interpretation is that being a native signer goes beyond timing of exposure, determining the quality and quantity of the input: native signers are likely to be the only population which is exposed in a natural environment to an fully fledged input, which might be lacking in school environments, where non-native signers are usually exposed to sign language. It is thus likely that the better performance of native signers is related to this qualitative and quantitative difference in the input received.

While this is certainly true, it cannot be the whole story. First of all, keep in mind that even Deaf parents are not a uniform class, and many might have themselves been exposed to sign language at a late period in life, thus providing an input that is not qualitatively different, at least as far as pure linguistic properties are concerned, from the input the general population is exposed to (cf. Lillo-Martin, 2021, and Singleton and Newport, 2004). Second, this "qualitative/quantitative" explanation would not extend to other findings pointing at a privilege of those children who are very early exposed to language as opposed to early exposed ones, no matter the family environment they are immersed in.

Friedmann and Szterman (2006) studied the competence in Hebrew of a group of hearing-impaired Hebrew-speaking (hence orally trained) children, all growing in hearing families under very similar circumstances. They found that individual performance in comprehension of long-distance dependencies in Hebrew was strongly correlated with the age of intervention: only children who received hearing aids before the age of 18 months performed well in the comprehension tasks. No other factors, such as the degree of hearing loss or the type of hearing device, significantly affected their performance. These findings indicate that something critical happens between birth and 1,5 years of age for syntax: in other words, they suggest that the critical period for first language syntactic competence is very early.

Friedmann and Rusou (2015) discuss the important issue of the effect of AoE in syntactic competences in a review paper, where they underline that most of the studies of a critical period for language acquisition test the acquisition of a second language, when one language has already been acquired. They suggest that a critical period for acquiring a first language is crucially different and earlier in time, and that for the acquisition of syntax it is the first year of life. While these results were only available until now with respect to spoken language inputs, the SIGN-HUB tests' contribution confirms the existence of this critical threshold also for sign language, which is not surprising considering that sign languages are natural languages just as any other, governed by the same bioprogram.

Be that as it may, this conclusion has important practical consequences that should be underlined in the most explicit way. Whether hearing aided or not, in order to guarantee unhindered language acquisition, deaf children should be exposed to sign language as early as possible, ideally from birth.

But Maybe Not Too Different

A question that we have not yet discussed is whether the lower performance that was captured in non-native signers is due to a competence gap (non-native signers have developed a different grammar) or to a performance gap (the resources necessary for computation are scarcer in non-native signers but the internal grammar is the same). Take the Subject Advantage in long-distance dependencies. The SIGN-HUB data show that this effect is stronger in non-native than in native signers. In the acquisition literature, the fact that the Subject Advantage in relative clauses and *wh*-questions gets reduced with age in simple picture matching tasks has been interpreted in terms of lower computational resources in young children. A similar explanation might be adopted here. Comprehending a first language acquired with a delay involves a bigger effort and this emerges in complex tasks. It was also noticed in Aristodemo et al. (in press) that a co-factor determining the particularly low performance of LSC non-native signers in the role shift comprehension task is the fact that in LSC stimuli non-manual markers were relatively subtle and might have not been noticed by non-native signers. This as well goes in the direction of a performance account.

If this were all that was found, we could conclude that native signers are different in that their performance is not affected by scarcity of resources, and they are more reliable as a source of linguistic information because their performance more directly reflects their grammatical competence. However, if we take a closer look at LSC for the relative clauses task, the picture appears to be different. In this language it was found that the Subject Advantage is so strong as to take the shape of a categorical difference between the grammar of native signers and that of non-native signers, who systematically misunderstand ORCs. The overall results suggest that while native signers have both SRCs and ORCs in their grammar, non-native signers do not allow ORCs at all in LSC. The fact that different varieties of languages realize different steps of the Accessibility Hierarchy of Keenan and Comrie (1977), which states that subject positions are more accessible than object positions in relativization, should not come as a surprise given the exceptional circumstances of access to language experienced by a large part of the deaf population. In fact, this finding, which replicates language internally the conclusion based on the typological literature, appears as an extreme case of AoE effect (Hauser et al., 2021).

If this is true, however, the question of the reliability of native signers gets partially reversed: if they sign a qualitatively different language, that is indeed a tight minority language within the community of signers, how can we capitalize on their language for description, pedagogical tools, standardization procedures, or language assessment? As for the latter, these findings advocate for the development of specific baselines at least distinguishing native from non-native signers. As for language description and its practical uses, the findings of the SIGN-HUB tests suggest that the common practice of relying exclusively on native signers should be complemented with a careful consideration of possible variations in different populations, crucially related to AoE.

CONCLUSION

In this paper we provide an overview of the morpho-syntactic SIGN-HUB assessment tests, a large-scale cross-linguistic study investigating comprehension across different sign languages and different syntactic phenomena to shed light on the notion of native signer. By relying on the same criteria to define native, early and late signers, the SIGN-HUB tests were able to provide new evidence that being exposed to sign language from birth has a permanent effect on language competence. In the syntactic tasks that were administered, native signers significantly outperformed non-native signers in a consistent way in most of the conditions tested.

While these results confirm that native signers perform differently from non-native signers, early learners included, they also suggest that at least for some phenomena and for some languages (and in particular for relative clauses in LSC) non-native learners develop a grammar that is significantly and qualitatively different from that of native signers. Overall, these results reaffirm the importance of native signers within the signing community, but also challenge the generalized use of the notion of native speaker/signer as the baseline for language description and language assessment. This is a crucial point when assessing clinical populations and should be considered through the life span, given that a delay in language exposure during childhood has permanent effects also in adulthood.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: for Hauser et al. (in press) https://osf.io/paj9n/?view_only=c9eaff3ba5a541cf9829a7de59a82e56; for Cecchetto et al. (2021) https://osf.io/g5cm9/; for Aristodemo et al. (in press) https://osf.io/emp6g/; for Hauser et al. (2021) https://osf.io/5bdu2/.

ETHICS STATEMENT

Ethical review and approval, and written informed consent, were not required for the current study in accordance with the local legislation and institutional requirements. As for the SIGN-HUB tests discussed in this article, they were reviewed and approved by for France (CERES, IRB n°. 20163400001072), for Italy (Milan Bicocca Ethical Committee, prot. n°. 0019845/16), for Spain (Parc de Salut MAR – Clinical Research Ethics Committee, prot. n°. 2016/6715/I). Participants provided their written informed consent to participate in these studies. Please refer to each specific paper for further details.

AUTHOR CONTRIBUTIONS

All authors contributed to the conceptualization of the manuscript. GZ wrote the first draft of the manuscript with the help of CD. GZ revised the manuscript with the help of CD and BG. All authors approved the final manuscript. CD, CC, and JQ supervised the project and acquired the funding.

FUNDING

The research summarized in this manuscript is part of the SIGN-HUB project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 693349. This publication has been funded in a joint effort by Laboratoire de Linguistique Formelle and the Labex Empirical Foundations of Language (Université Paris Cité, CNRS).

REFERENCES

Aristodemo, V., and Friedmann, N. (2019). *COGONEOUTLSF- Cognitive Odd One Out Test for LSF. Handle to be Attributed*. Available online at: https://www.signhub.eu/assessment/lsf

Aristodemo, V., Giustolisi, B., Checchetto, C., and Donati, C. (2020). "Comprehension of verb directionality in LIS and LSF," in *Proceedings of the EXLING 2020 11th International Conference of Experimental Linguistics*, ed. A. Botinis (Athens), ExLing Society. doi: 10.36505/ExLing-2020/11/0008/000423

Aristodemo, V., Giustolisi, B., Zorzi, G., Gras, D., Hauser, C., Sala, R., Sánchez Amat, J., et al. (in press). On the nature of role shift: insights from a comprehension study in different populations of LIS, LSC, and LSF signers. *Nat. Lang. Ling. Theory*

Aristodemo, V., Hauser, C., Cecchetto, C., Friedmann, N., and Donati, C. (2019). *SYNTCQLSF-Content Questions Comprehension Task in LSF. Handle to be Attributed*. Available online at: https://www.signhub.eu/assessment/lsf

Bauman, H.-D. L., and Murray, J. J. (2017). "Sign languages," in *The Oxford Handbook of Language and Society*, eds O. García, N. Flores, and M. Spotti (Oxford: Oxford University Press), 243–260.

Boudreault, P., and Mayberry, R. I. (2006). Grammatical processing in American Sign Language: age of first-language acquisition effects in relation to syntactic structure. *Lang. Cogn. Process.* 21, 608–635. doi: 10.1080/01690960500139363

Branchini, C., and Donati, C. (2009). "Relatively different: Italian sign language relative clauses," in *Correlatives Cross-Linguistically*, ed. A. Liptak (Amsterdam: Benjamins), 157–191. doi: 10.1075/lfab.1.07bra

Carreiras, M., Gutiérrez-Sigut, E., Baquero, S., and Corina, D. (2008). Lexical processing in Spanish sign language (LSE). *J. Mem. Lang.* 58, 100–122.

Cecchetto, C., Checchetto, A., Giustolisi, B., and Santoro, M. (2021). Age of exposure and subject/object asymmetries when wh-movement goes rightward: the case of LIS interrogatives. *Sign Lang. Ling.*

Cecchetto, C., Geraci, C., and Zucchi, S. (2006). Strategies of relativization in Italian sign language. *Nat. Lang. Ling. Theory* 24, 945–975.

Cecchetto, C., Geraci, C., and Zucchi, S. (2009). Another way to mark syntactic dependencies: the case for right-peripheral specifiers in Sign Language. *Language* 85, 278–320. doi: 10.1353/lan.0.0114

Checchetto, A., Santoro, M., Giustolisi, B., Friedmann, N., Donati, C., and Cecchetto, C. (2019). *SYNTCQLIS-Content Questions Comprehension Task in LIS. Handle to be Attributed*. Available online at: https://www.sign-hub.eu/assessment/lis

Chen Pichler, D., and Lillo-Martin, D. (2018). "Hearing parents and deaf children learning a sign language together," in *Proceedings of the Talk at the 4th International Congress on Family-Centered Early Intervention for Children who are Deaf or Hard-of-Hearing*, Bad Ischl.

Cormier, K., Schembri, A., Vinson, D., and Orfanidou, E. (2012). First language acquisition differs from second language acquisition in prelingually deaf signers: evidence from sensitivity to grammaticality judgement in British Sign Language. *Cognition* 124, 50–65. doi: 10.1016/j.cognition.2012.04.003

Costello, B., Fernandez, F., and Landa, A. (2008). "The non-(existent) native signer: sign language research in a small deaf population," in *Proceedings of the Theoretical Issues in Sign Language Research (TISLR) 9 Conference* 2006, eds B. Costello, J. Fernández, and A. Landa (Florianopolis: Editora Arara Azul).

Diessel, H., and Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language* 81, 882–906. doi: 10.1353/lan.2005.0169

Dye, M. W., and Shih, S. I. (2009). Phonological priming in British sign language. *Lab. Phonol.* 8, 241–264.

Emmorey, K., Bellugi, U., Friederici, A., and Horn, P. (1995). Effects of age of acquisition on grammatical sensitivity: evidence from on-line and of-line tasks. *Appl. Psycholing.* 16, 1–23.

Frazier, L. (1987). "Sentence processing: a tutorial review," in *Attention and Performance 12: The Psychology of Reading*, ed. M. Coltheart (Mahwah NJ: Lawrence Erlbaum Associates), 559–586. doi: 10.1007/BF01708572

Freel, B. L., Clark, D., Anderson, M., Gilbert, G., Musyoka, M., and Hauser, P. (2011). Deaf individuals' bilingual abilities: American sign language proficiency, reading skills, and family characteristics. *Psychology* 2, 18–23. doi: 10.4236/psych.2011.21003

Friedmann, N., Belletti, A., and Rizzi, L. (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88.

Friedmann, N., and Rusou, D. (2015). Critical period for first language: the crucial role of language input during the first year of life. *Curr. Opin. Neurobiol.* 35, 27–34. doi: 10.1016/j.conb.2015.06.003

Friedmann, N., and Szterman, R. (2006). Syntactic movement in orally trained children with hearing impairment. *J. Deaf Stud. Deaf Educ.* 11, 56–75. doi: 10.1093/deafed/enj002

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. *Image Lang. Brain* 2000, 95–126. doi: 10.1016/j.cognition.2006.09.011

Giustolisi, B., and Friedmann, N. (2019). *COGONEOUTLIS – Cognitive Odd One Out Test for LIS. Handle to be Attributed*. Available online at: https://www.signhub.eu/assessment/lis

Giustolisi, B., Sala, R., Friedmann, N., Donati, C., and Cecchetto, C. (2019). *SYNTRELLIS – Relative Clauses Comprehension Task in LIS. Handle to be Attributed*. Available online at: https://www.sign-hub.eu/assessment/lis

Hall, M., Wyatte, C. H., and Caselli, N. K. (2019). Deaf children need language, not (just) speech. *First Lang.* 39, 367–395. doi: 10.1177/0142723719834102

Hauser, C. (2019). *Subordination in LSF: Nominal and Sentential Embedding.* Doctoral dissertation. Paris: Université de Paris.

Hauser, C., Cecchetto, C., Friedmann, N., and Donati, C. (2019). *SYNTRELLSF-Relative Clauses Comprehension Task in LSF. Handle to be Attributed*. Available online at: https://www.sign-hub.eu/assessment/lsf

Hauser, C., Zorzi, G., Aristodemo, V., Giustolisi, B., Gras, D., Sala, R., et al. (2021). Asymmetries in relative clause comprehension in three European sign languages. *Glossa J. Gen. Ling.* 6, 1–36. doi: 10.5334/gjgl.1454

Hauser, C., Aristodemo, V., and Donati, C. (in press). Subject advantage in covert dependencies: the case of wh-questions comprehension in French sign language. *Syntax*

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language* 75, 244–285. doi: 10.2307/417261

Humphries, T., Kushalnaga, P., Mathur, G., Jo Napoli, D., Padden, C., Rathmann, C., et al. (2016). Language choices for deaf infants: advice for parents regarding sign languages. *Clin. Pediatr.* 55, 513–517. doi: 10.1177/0009922815616891

Jaeger, H. (2019). Not like them": new Signers' narrative constructions of the authentic signer. *Ampersand* 6:100048. doi: 10.1016/j.amper.2019.100048

Jaffe, A. (2015). Defining the new speaker: theoretical perspectives and learner trajectories. *Int. J. Sociol. Lang.* 231, 21–44. doi: 10.1515/ijsl-2014-0030

Johnston, T. (2006). W(h)ither the deaf community? Population, genetics, and the future of Australian sign language. *Sign Lang. Stud.* 2, 137–173. doi: 10.1353/aad.2004.0004

Joint Committee on Infant Hearing (2019). Year 2019 position statement: principles and guidelines for early hearing detection and intervention programs. *J. Early Hear. Detect. Interv*. 4, 1–44. doi: 10.15142/fptk-b748

Keenan, E. L., and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Ling. Inquiry* 8, 63–99.

King, J., and Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30, 580–602. doi: 10.1016/0749-596x(91)90027-h

Lillo-Martin, D. (2021). "Chomsky and signed languages," in *Wiley-Blackwell Companion to Chomsky*, eds N. Allott, T. Lohndal, and G. Rey (Hoboken, NJ: Wiley). doi: 10.1002/9781119598732.ch22

Mak, W. M., Vonk, W., and Schriefers, H. (2006). Animacy in processing relative clauses: the hikers that rocks crush. *J. Mem. Lang.* 54, 466–490.

Mathur, G., and Rathmann, C. (2006). "Variability in verbal agreement forms across four signed languages," in *Laboratory Phonology VIII*, eds L. Goldstein, D. Whalen, and C. T. Best (Berlin: Mouton), 287–314.

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: the case of American Sign Language. *J. Speech Lang. Hear. Res.* 36, 1258–1270. doi: 10.1044/jshr.3606.1258

Mayberry, R. I. (2010). "Early language acquisition and adult language ability: what sign language reveals about the critical period for language," in *Oxford Handbook of Deaf Studies, Language, and Education*, eds M. Marschark and P. Spencer (Oxford: Oxford University Press), 281–291.

Mayberry, R. I., and Kluender, R. (2018). Rethinking the critical period for language: new insights into an old question from American Sign Language. *Biling. Lang. Cogn.* 21, 886–905. doi: 10.1017/S1366728917000724

Mayberry, R., Lock, E., and Kazmi, H. (2002). Development: linguistic ability and early language exposure. *Nature* 417:38. doi: 10.1038/417038a

Mitchell, R. E., and Karchmer, M. A. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in the United States. *Sign Lang. Stud.* 4, 138–163.

Mosella Sanz, M. (2012). *Les Construccions Relatives en Llengua de Signes Catalana (LSC). Doctoral Dissertation*. Barcelona: University of Barcelona.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., and Lee, R. (2000). *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, MA: MIT Press.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: studies of the acquisition of American Sign Language. *Lang. Sci.* 10, 147–172. doi: 10.1016/0388-0001(88)90010-1

O'Rourke, B., Pujolar, J., and Ramallo, F. (2015). New speakers of minority languages: the challenging opportunity. *Int. J. Sociol. Lang.* 231, 1–20. doi: 10.1177/1524839918761865

Quer, J. (2002). *Negative Operators in Catalan Sign Language*. Barcelona: ICREA & Universitat de Barcelona.

Quer, J. (2011). Reporting and quoting in signed discourse. *Understanding Quot.* 7:277.

Quer, J., Rondoni, E. M., Barberà, G., Frigola, S., Aliada, D., and Boronat, J. (2005). *Gramàtica Bàsica LSC*. Barcelona: FESOCA, DOMAD.

Quer, J., and Steinbach, M. (2019). Handling sign language data: the impact of modality. *Front. Psychol.* 10:483. doi: 10.3389/fpsyg.2019.00483

Singleton, J. L., and Newport, E. (2004). When learners surpass their models: the acquisition of American sign language from inconsistent input. *Cogn. Psychol.* 49, 370–407. doi: 10.1016/j.cogpsych.2004.05.001

Sorace, A. (2021). Monolinguals: a species facing extinction? *ExLing Tutorial* 30:2021.

Tupi, E. (2019). *Sign Language Rights in the Framework of the Council of Europe and Its Member States*. Helsinki: Ministry for Foreign Affairs of Finland.

Van Valin, R. Jr., and Wilkins, D. (1996). "The case for 'effector': case roles, agents, and agency revisited," in *Grammatical Constructions: Their Form and Meaning*, eds M. Shibatani and D. Thompson (Oxford: Clarendon Press), 289–322.

Woll, B. (2018). The consequences of very late exposure to BSL as an L1. *Biling. Lang. Cogn.* 21, 936–937. doi: 10.1017/s1366728918000238

Zorzi, G., Sánchez Amat, J., Cecchetto, C., Donati, C., Friedmann, N., and Quer, J. (2019a). *SYNTRELLSC – Relative Clauses Comprehension Task in LSC. Handle to be Attributed*. Available online at: https://www.signhub.eu/assessment/lsc

Zorzi, G., Sánchez Amat, J., Cecchetto, C., Friedmann, N., Donati, C., and Quer, J. (2019b). SYNTCQLSC – Content Questions Comprehension Task in LSC. Handle to be Attributed. Available online at: https://www.signhub.eu/assessment/lsc

Zorzi, G., Sánchez Amat, J., and Friedmann, N. (2019c). COGONEOUTLSC – Cognitive Odd One Out Test for LSC. Handle to be Attributed. Available online at: https://www.sign-hub.eu/assessment/lsc

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

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
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership